specialbulletin från

PEDAGOGISK-
PSYKOLOGISKA
INSTITUTIONEN

LÄRARHÖGSKOLAN
MALMÖ

testkonstruktion och testdata

Bierschenk, I.:

KONSTRUKTION AV ETT REGELSYSTEM FÖR EN DATORBASERAD INNEHÅLLSANALYS AV INTERVJUTEXT: PRELIMINÄRMANUAL OCH NÅGRA UTPRÖVNINGSRESULTAT

Nr 25 November 1974

KONSTRUKTION AV ETT REGELSYSTEM FÖR EN DATORBASERAD INNEHÅLLSANALYS AV INTERVJUTEXT: PRELIMINÄRMANUAL OCH NÅGRA UTPRÖVNINGSRESULTAT

Inger Bierschenk

Bierschenk, I. Konstruktion av ett regelsystem för en datorbaserad innehållsanalys av intervjutext: Preliminärmanual och några utprövningsresultat. /Construction of rules for a computer-based content analysis of interview texts: A preliminary manual and some evaluation data./ Testkonstruktion och testdata (Malmö, Sweden: School of Education), Nr 25, 1974.

Denna rapport innehåller en beskrivning av en teknik för en datorbaserad innehållsanalys av intervjutext. I rapporten redovisas några steg i utvecklingen av ett regelsystem för kodning av intervjutext. Utprövningen av ett preliminärt regelsystem beskrivs och några utprövningsdata presenteras. Dessutom redovisas detaljerat hur kontrollen av stansningen och stansunderlaget utförts för att erhålla ett felfritt material. En språklig diskussion föregår presentationen av det hittills utförda arbetet.

Nyckelord: psycholinguistics, concept formation, interviews, data processing, content analysis.

INNEHÅLL

DEL I. EN INTRODUCERANDE DISKUSSION OM SPRÅK OCH SPRÅKANALYS

1. ANALYSENS PLATS OCH SYFTE INOM ETT FORSKNINGSPROJEKT

Vid Lärarhögskolan i Malmö pågår på pedagogisk-psykologiska institutionen ett forskningsprojekt kallat Skolpedagogiska sökstrategier (SÖK), som finansieras av Skolöverstyrelsen. Projektets huvudsakliga syften kan sägas vara att kartlägga hur forskare på Sveriges pedagogiska institutioner ser på (1) problemperception och problemformulering och (2) informationssökning och -spridning. Den hittills genomförda datainsamlingen har skett genom intervjuer. 40 forskare vid pedagogiska institutioner i Sverige har uttagits slumpmässigt och intervjuats vårterminen 1973. Intervjuerna spelades in på ljudband. Förutsättningarna för och genomförandet av datainsamlingen redogörs för i B. Bierschenk (1974 a).

Det material som skall utgöra grunden för utvecklingen av den datorbaserade analysmetoden är det bandade materialet. Detta har nedskrivits och omfattar ca 4000 sidor text. Nedteckningen har inte skett med beaktande av fonologiska transkriberingsprinciper, men i övrigt så autentiskt som möjligt. Alla hörbara yttranden har skrivits ner, vilket innebär att talarens felsägningar, korrigeringar och ofullbordade meningar finns med. Materialets karaktär och omfång gör det nödvändigt att dels

(1) utarbeta regler för hur forskarnas yttranden skall brytas ner i hanterliga enheter, dels

(2) möjliggöra bearbetning och tolkning med hjälp av dator.

Ett preliminärt försök till regelutveckling har utförts (I. Bierschenk, 1974 a och b). Provkodningar och reliabilitetsbedömningar har genomförts och resultaten har presenterats (Berg, 1974 a och b; I. Bierschenk, 1974 c). Enligt de regler som hittills föreligger är ca 10 % av materialet bearbetat och kodat och föreligger färdigt för maskinell bearbetning. Några datorprogram för enklare bearbetningar finns redan.

Ett viktigt led i arbetet har också varit kontroller av det kodade materialets skick efter det att det stansats på hålkort. Kontrollstansning och granskning av feltyper samt rättning av hela materialet har därför genomförts och rapporterats (I. Bierschenk, 1974 d). Samtliga led i detta utvecklingsarbete redovisas i denna rapport. Innan detta sker kan det emellertid vara lämpligt att kortfattat beskriva analysens plats i ett vidare perspektiv.

2. DATORBASERAD INNEHÅLLSANALYS

2.1 Några förutsättningar

I den beteendevetenskapliga forskningen används i stor utsträckning innehållsanalys, när man vill utvinna information ur verbaliserade svar. Det finns emellertid ingen innehållsanalysmetod som gäller generellt för allt verbalt material. Till följd av att forskare inom olika discipliner sedan länge varit bundna till manuella analyser, har det utvecklats lika många innehållsanalystekniker som det funnits forskare som använt innehållsanalysmetoder. Dessutom är innehåll i en text inte något givet, utan det blir definierat först genom den teknik som använts i varje enskilt fall. Först nu tycks det finnas tecken på att denna metod kan bli mera enhetlig och således mera generell, i och med att datorer kommit till användning.

Utvecklingen av en datorbaserad innehållsanalysmetod som lämpar sig för varierande syften inom de beteendevetenskapliga ämnena är i hög grad önskvärd. Den beteendevetenskapliga forskningen styrs vid informationsinsamlingen i alltför stor utsträckning av formaliserade frågeformulär. De data som på så sätt insamlats är relativt lätta att hantera jämfört med verbaliserade svar, men metoden innebär i många fall att risken för ofullständig information är stor. De forskare som inte vill ta sådana risker och därför tillåter öppna svar står genast inför problemet att bearbeta text, dvs att analysera data med i princip obegränsade tolkningsmöjligheter. Ett välkänt faktum är också att en manuell bearbetning av text är mycket tidsödande och arbetskrävande, eftersom den information man är ute efter sällan finns väl samlad. Vid organisering och strukturering av sådan information torde datorn utgöra ett viktigt hjälpmedel. En datorbaserad innehållsanalys är oftast ofrånkomlig i de fall där forskare önskar utföra sofistikerade analyser och där de vill pröva olika teorier och modeller på samma grundmaterial utan att för den skull behöva manuellt bygga om komplexa kategorisystem och omkoda stora mängder av informationsenheter.

2.2 Anknytning till annan forskning

Förutsättningen för att redan utvecklade metoder skall kunna tillämpas är ofta knuten till materialets utseende. Otaliga försök till innehållsanalytisk tolkning av text har gjorts och pågår framför allt i USA. Ett av de försök som ofta refereras i litteraturen är Osgoods "Evaluative Assertion Analysis" (Osgood et al, 1956). Osgood presenterar en metod för särskiljande av s k attitydobjekt från s k common meaning. Attitydobjekten är i första hand substantiv, som värderas med hjälp av skalerade predikativa påståenden

och attribut som är knutna till dessa attitydobjekt i texten. På så sätt anses metoden mäta attityd/värdering av vissa företeelser. Metoden är för våra syften inte tillräcklig, då den inte använder hela texten och dessutom inte är datorbaserad.

En vidareutveckling, en datorbaserad analogi till Osgoods metod, presenteras av Holsti (1969). Medan den förra metoden använder substantiv och adjektiv, för den senare in även verben som betydelsefulla för värderingen av objekt. Man tar också hänsyn till satstema ("modes of expression"), t ex negation, modalitet. Därmed indikerar Holsti att syntaktisk kodning är ofrånkomlig. Det är omöjligt att skilja ut enstaka "ord" och analysera dem oberoende av den omgivande texten. Syntaktisk struktur innebär ordning mellan ord i satser, ordning mellan satser i meningar och ordning mellan meningar i text. En analys av hel text måste följaktligen ta hänsyn till samtliga dessa beroendeförhållanden. Denna beroendestruktur som kännetecknar text är dock mer eller mindre tydlig beroende på situationen i vilken texten produceras (tal - skrift, interpersonella relationer mellan meddelare och mottagare, avsikten med texten, "producentens" psykologiska tillstånd, etc).

De i litteraturen redovisade analysförsöken är baserade på skriven text av typen politiska tal, nyhetsartiklar, etc. Sådana texter är genomtänkta alster, som är avsedda att förmedlas till en stor publik. Vårt material är talad text, som i yttrandeögonblicket endast varit avsedd för en person, intervjuaren. Jämfört med skriven text innehåller en sådan "spontan" text en hel del information om den talandes värderingar och ståndpunktstaganden, som vi tror inte skulle framkomma om individernas åsikter hade framförts skriftligt och formellt. T ex antar vi att ordvalet är personligare än vad som skulle "tillåtas" i skrift. Vissa talspråkliga former som uttryck för modifiering av yttranden och som uttryck för inställningen till den specifika interpersonella kontaktsituationen vid intervjun förekommer antagligen i hög grad. Mycken sådan information, som psykologiskt kan tolkas (rädsla, avståndstagande, frimodighet, m m), går förlorad i skrift, där den officiella karaktären kräver annorlunda och precisare uttrycksmedel.

3. SAMVERKAN MELLAN LINGVISTIK OCH PSYKOLOGI

Användning av språk är en psykologisk och fysiologisk process, som innebär överföring av tankar från en individ till en annan via tal och skrift (eller gester). Språket är medlet som vi använder för att uttrycka vår erfarenhet av processer, personer, objekt, abstraktioner, kvaliteter, tillstånd och relationer i den värld som omger oss och finns inom oss. En kartläggning av relationerna mellan erfarenhetsvärlden och språket kan kallas psykolingvistik.

Erfarenheter kan sägas bestå av relationer mellan begrepp (conceptualizations). Schank (1972, s 556) skriver:

"The important point is that underlying every sentence in a language there exists at least one conceptualization. Conceptualizations may relate to other conceptualizations by nesting or other specified relationships, so it is possible for a sentence in a language to be the realization of many conceptualizations at one time. This is like saying that one sentence can express many complete ideas and the relation of those ideas."

Språket är som ett isberg, där endast ytan är synlig för omedelbar observation och där större delen, idénätverket, endast långsamt kan utforskas i hela dess omfång. Många metoder har utvecklats i syfte att studera underliggande betydelser i yttranden under beteckningen 'innehållsanalys'. Skillnaden mellan lingvisten och innehållsanalytikern kan sägas bestå i att innehållsanalytikern främst använder statistiska metoder, men, som Pool (1961, s 324) anför:

"the barriers between linguistics and content analysis seem likely to give way, and it seems probable at least that content analysis methods will be increasingly enriched by knowledge of the findings of linguistic research."

Ett av lingvistens viktigaste bidrag till innehållsanalysforskningen är att fastställa de karakteristika i språket som är allmänna och vilka som är specifika för en individ eller situation. De är de senare som är indikatorer för psykolingvisten att arbeta med.

Språk är ett uttryck för process: handlingar, händelser, tillstånd och relationer och därtill knutna personer, objekt och abstraktioner. Denna process uttrycks genom satser i en viss ordningsföljd. Processen själv representeras av verbet. Deltagare i processen är t ex personer och objekt. Dessa spelar rollen av agent, mål, etc (se Halliday, 1970). Detta "rollspel" i förhållande till verbet kallas transitivitet. Det verb som en individ väljer att uttrycka en process med är med Schanks term ett concept. Beroende på verbet och på rollspelets parter bildas en conceptualization (begreppsrelation). Den språkliga analys vi avser utveckla utgår från satser och kan därför sägas vara syntaktisk. Men eftersom vi vill ta hänsyn till transitiviteten, kan vi inte alltid bilda strikt grammatiska kategorier utan snarare logiska.

Schank (1972, s 566) skriver:

"In attempting to uncover the actual conceptualization underlying a sentence, we must recognize that a sentence is often more than its component parts. In fact, a dialogue is usually based on the information that is left out of a sentence but is predicted by the conceptual rules."

Detta betyder att man vid analys av 'innehåll' i en text inte kan arbeta med enstaka 'ord'. Orden är representanter för hela begreppsrelationer. När vi yttrar en tanke, gör vi detta så ekonomiskt som situationen fordrar. I en dialog mellan individer som kan anses ha likartade nätverk av begreppsrelationer (t ex två forskare inom samma område) kan individerna lätt kommunicera i och med att de använder sig av sådan verbal representation att samma begreppsrelation framkallas hos meddelare och mottagare. Därför är det troligt att språket i en sådan situation, dvs alla yttranden tillsammans (kontext) består av en "ekonomisk" representation, rik på underliggande information. Hela begreppsrelationer behöver inte vara uttalade.

Dessutom måste det tas hänsyn till den ekonomi som ligger i språket självt, beroende på förhållandet mellan beroende och oberoende begrepp. De oberoende är de som ensamma kan framkalla begreppsrelationer. Schank kallar dessa för "picture producer". Det är alltså olika slag av substantiv och verb. De beroende är de som endast tillsammans med de oberoende kan bilda conceptualizations, dvs bestämningar, attribut (adjektiv) och adverb.

Ex. A: "Har du ett eller flera projekt på gång?"

B: "Flera."

I svaret ligger underförstått "projekt". Men eftersom både A och B vet detta i denna situation, behöver B inte säga mer. Hela begreppsrelationen i B:s svar är: "Jag har flera projekt på gång".

I en innehållsanalys på satsanalytisk grund blir satsen minsta relationsenhet. Men satser ingår som delar i större enheter, varför det måste tas hänsyn till relationerna mellan satserna. Den inneboende ekonomin i satslänkning kan exemplifieras med relationsordet 'som', som är ett beroende begrepp, som endast blir tolkbart genom supplering av det oberoende begrepp det i varje enskilt yttrande innefattar. I en ordinär syntaktisk analys skulle vi säga att en sats som

"Dom sökte upp forskningsgruppen"

är fullständig. Dom är subjekt, sökte upp är predikat och forskningsgruppen är objekt. Men för en syntaktisk-begreppsbaserad analys är den ofullständig. Länken till föregående yttrande saknas, dvs vi vet inte vem 'Dom' är. Vi vet inte heller vilken forskningsgrupp som avses. Att ta hänsyn till begrepp innefattar också transitivitetsbestämning. 'Dom' agerar i en bestämd situa-

tion med en bestämd intention, nämligen att komma i kontakt med en viss forskningsgrupp. Vi måste analysera såväl den agerande som handlingens mål: "Dom (forskarstuderande i pedagogik från Malmö) sökte upp forskningsgruppen (för Skrivsyntax i Lund)". Denna kodade sats skulle ur grammatisk synpunkt vara fullständig för vår analys. Det återstår att bestämma rollspelets betydelse i processen. Vad tycker talaren om forskarstuderande i pedagogik i Malmö egentligen? Det beror på vad han tycker om forskningsgruppen för Skrivsyntax i Lund. En negativ inställning till forskningsgruppen kan medföra att den "neutrala" handlingen att "söka upp" här uppfattas som negativ. Vad varje enskild sats (process) har för innehåll kan bara avgöras efter analys av vilka begreppsrelationer som talaren "döljer" i den övriga texten angående de kontrahenter som är inblandade i den yttrade processen. Det är i detta tolkningsskede som datorn skall hjälpa oss vidare. Schank (1972, s. 627) påpekar:

"In order to deal with thoughts on a machine it is necessary to characterize them in some way. That is, we must extract the inherent ideas from the linguistic input and characterize these ideas in some fashion so to be able to use them. It is the use of these ideas that has been sorely neglected by linguists, yet it is precisely the use of ideas that is the communication process."

En relation mellan begrepp måste alltid innefatta en process, som uttrycker handling, händelse eller tillstånd. Varje gång vi konfronteras med ett begrepp, som för oss är känt, framkallas inom oss bilden av en process. Vad är det som gör att jag är rädd för åska? Ordet 'åska' är för mig känt som någonting som kan agera mot mig genom att blixten slår ner och dödar mig. Jag framkallar bilden av en möjlig process som för mig vore negativ. Därmed säger jag att ordet 'åska' är negativt. En psykolingvistisk analys kan inte negligera den erfarenhetsvärld som den enskilde individen eller grupp av individer explicit eller implicit uttrycker genom sitt språk. Om man antar att en explicit värdering av personer, objekt och abstraktioner uttrycks genom bruket av attribut (adjektiv) och predikativa påståenden om dem, får vi också anta att bruket av verb implicit kan uttrycka detsamma. Personer, objekt och abstraktioner kan omedelbart och synligt bli modifierade (värderade) genom attribut (adjektiv), och eftersom adjektiv är beroende begrepp kan vi, även där det oberoende begreppet inte omedelbart yttras i nära anslutning, sluta oss till vilket oberoende begrepp som avsetts. Men i de fall där en process bestående av enbart oberoende begrepp yttras, kan en värdering döljas i själva processen, dvs i beskrivningen av de syntaktiska rollrelationerna. Att utröna huruvida värderingar av personer, objekt och abstraktioner inte bara explicit uttrycks genom beroende begrepp utan även implicit genom individens val att framställa dem som parter i

en process är en angelägen uppgift för såväl lingvister som psykologer. Att utveckla metoder för hur en manuell respektive datorbaserad analys skall ske är både tids- och arbetskrävande men torde kunna vara ett steg på vägen till vidgade psykolingvistiska studier.

3.1 Analys av begrepp via dator

De teorier som ovan presenterats, såväl Schanks psykologiska resonemang som Hallidays språkligt förankrade, har projektet funnit mycket intressanta. Vi har däremot inte haft dem till utgångspunkt när regelutvecklingsarbetet startade, eftersom vi inte förrän några månader senare kom i kontakt med dem. De förefaller emellertid båda vara möjliga att integrera i den fortsatta utvecklingen, särskilt som många av dessa tankar implicit finns uttryckta i vårt sätt att angripa problematiken, t ex suppleringsförfarandet och koncentrationen kring verbet som bärare av en process. På detta stadium kan vi dock ännu inte avgöra i vad mån analysen kommer att gå i en sådan riktning att det blir möjligt att analysera sådana tankestrukturer som Schank diskuterar, även om det är vår förhoppning. Vår utgångspunkt har i första hand varit paradigmet agent-handling-objekt-mål, som Holsti presenterat, vilket ligger mycket nära Hallidays rollrelationer. Man kunde också ha anknutit till Fillmore (1968), som framlägger en kasus-teori där de olika kasus representerar ett liknande rollförhållande. Ett försök att skriva en kasus-grammatik för svenskan har gjorts av Brodda (1973), där han anför prepositioner som kasus-signalerare. Men det påpekas också att sådana kasus (s k "inre") som representerar logiska relationer inom satser inte nödvändigtvis markeras genom prepositioner. Dessutom har vi s k "yttre" kasus (tids- och rumsuttryck), där prepositionerna är desamma som i övriga fall (jfr Brodda, 1973, s 58). Man kan således inte bygga regler med hjälp av prepositioner. Att under sådana förhållanden bygga på en kasus-modell är mycket svårt, om man vill undvika tolkningsolikheter mellan flera kodare. En sådan kodning skulle dessutom innebära en semantisk analys, där "mening" eller "innehåll" a priori bestäms. Våra "roller" kodas i förhållande till verbet, för att riktningen i en handling skall framträda. Mellan de begrepp som kodas i relationen med verbet gör vi ingen skillnad (människa, icke-personella objekt, o s v). Det betyder att vi inte, som semantikerna, specificerar alla ingångar ("entries") i form av lexikal bestämning. Den differentieringen inom koderna (kategorierna) lämnas öppen tills en datorbearbetning har visat vilka grupperingar och hierarkier som skall betraktas som relevanta.

Vad som skiljer vår analys från de ovan nämnda teoretiska ansatserna är att vi avser empiriskt förankra vårt "innehåll". Först med ett större textmaterial som detta blir det möjligt att pröva olika teorier.

Genom att en syntaktisk relation har bestämts, har vi skapat kategorier som skall bilda basen för den fortsatta analysen. En huvudsaklig frågeställning inom projektet är: Vilka värderingar har de intervjuade forskarna och hur värderas de handlingar och "roller" (agent-objekt) som inbegripes i forskningsprocessen? Våra kategorier (begrepp) bildar enheter i register, som byggs upp med hjälp av Osgood's semantiska differentialer. Beroende begrepp kommer att skaleras i de tre dimensionerna värdering, aktivitet och intensitet (styrka). Beroende på resultatet av denna skalering kommer våra första strukturer att framträda. Först här kan vår "datorbaserade innehållsanalys" börja, i och med att vi kan arbeta med empiriskt förankrade facetter. Vid konstruktionen avser vi att använda statistiska modeller för bestämning av vilka strukturer och nivåer som finns i materialet. En utförligare presentation av vår analysmetod (Analysis of Concepts by Data-processing, ANACONDA) ges i B. Bierschenk (1974 b).

Det arbete som hittills rapporterats är således en bas för den fortsatta utvecklingen. De datorbaserade analyserna kommer att bestämma om och på vilket sätt våra regler skall specificeras och förfinas.

Det hade varit lämpligast att presentera vår preliminärmanual här, eftersom de följande kapitlen innehåller begrepp från den. I sin första reviderade form från juni 1974 föreligger den som särstencil, varför jag av tekniska skäl i stället låter den avsluta rapporten som en egen enhet (del III).

DEL II. ETT REGELSYSTEMS UTVECKLING OCH UTPRÖVNING

Ett av stegen i utvecklingen av intervjun var fyra provintervjuer (se B. [illegible], 1979). Syftet med dessa var i första hand att vara vägledande för intervjuaren i den slutliga utformningen av intervjun och [illegible] av intervjutekniken som datainsamlingsmetod. Då även provintervjuerna, som utfördes vid lärarhögskolan i Malmö, spelades in på band och skrevs ut, kunde de användas i det första skedet av regelsystemets utveckling, när vi skulle göra försök med nedbrytning av text.

4.1 De första stegen

De fyra provintervjuerna lästes igenom. Vi ville få en uppfattning om dels (1) materialets skick vad gäller utskriften och textens komplexitet ifråga om meningslängd och förekomst av ord för en läsare som inte varit med vid intervjutillfället, dels (2) hur vi skulle kunna attackera och behandla denna form av information. Osgoods och Holstis innehållsanalysmetoder (se kap 2) diskuterades med avsikten att pröva i vad mån dessa kunde överföras på svensk text och vårt material. Det stod tidigt klart att texterna måste förbehandlas, innan vi fick lämpliga analysenheter. Ett från början skrivet material är sedan strukturerat med avseende på meningsbyggnad och interpunktion. Här fanns alla yttranden med, såväl grammatiskt korrekta som felaktiga, ofärdiga, fragment osv. Vi kom fram till att betrakta den fullständiga meningen (sentence) som den mest relevanta analysenheten. Detta innebar samtidigt att vi skulle komma att röra oss med satsbegreppet. Vissa yttranden som var så fragmentariska att de inte kunde förstås i sammanhanget borde därför utgå för att vi skulle undvika tolkningsproblem.

När vi så hade försökt att dela in textstycken i fullständiga meningar, diskuterade vi de olika delarnas innehåll och kom fram till att knappast någon mening (eller sats) uppbar en tillräcklig information utan att vara beroende av det närmast föregående. Kompletteringar måste således göras för att en mening ensam skulle kunna bli förståelig ifall den som analyserade tagit den ur sitt sammanhang (kontext).

Dessa våra första försök utmynnade i en preliminär regelsamling för

1. segmentering, komplettering och strykning av text.

Detta förfaringssätt tillämpades in och diskuterades, innan vi påbörjade försöken att tolka de enskilda leden inom varje större analysenhet. Utgångspunkten här var Holstis paradigm agent-handling-objekt-mål. Dessa relationer tränade vi oss i att upptäcka och fastställa sedan.

4. REGELSYSTEMETS UTVECKLING

Ett av stegen i utvecklingen av intervjun var fyra provintervjuer (se B. Bierschenk, 1974). Syftet med dessa var i första hand att vara vägledande för intervjuaren i den slutliga utformningen av intervjun och hanteringen av intervjutekniken som datainsamlingsmetod. Då även provintervjuerna, som utfördes vid lärarhögskolan i Malmö, spelades in på ljudband och skrevs ut, kunde de användas i det första skedet av regelsystemets utveckling, när vi skulle göra försök med nedbrytning av text.

4.1 De första stegen

De fyra provintervjuerna lästes igenom. Vi ville få en uppfattning om dels (1) materialets skick vad gäller utskriften och textens komplexitet ifråga om meningslängd och förståelighet för en läsare som inte varit med vid intervjutillfället, dels (2) hur vi skulle kunna attackera och behandla denna form av information. Osgoods och Holstis innehållsanalysmetoder (se kap 2) diskuterades med avsikten att pröva i vad mån dessa kunde överföras på svensk text och vårt material. Det stod tidigt klart att texterna måste förbehandlas, innan vi fick lämpliga analysenheter. Ett från början skrivet material är redan strukturerat med avseende på meningsbyggnad och interpunktion. Här fanns alla yttranden med, såväl grammatiskt korrekta som felaktiga, korrigeringar, fragment osv. Vi kom fram till att betrakta den fullständiga meningen (sentence) som den mest relevanta analysenheten. Detta innebar samtidigt att vi skulle komma att röra oss med satsbegreppet. Vissa yttranden som var så fragmentariska att de inte kunde förstås i sammanhanget borde därför utgå för att vi skulle undvika tolkningsproblem.

När vi så hade försökt att dela in textstycken i fullständiga meningar, diskuterade vi de olika delarnas innehåll och kom fram till att knappast någon mening (eller sats) uppbar en alldeles egen information utan att vara beroende av det närmast föregående. Kompletteringar måste ideligen göras för att en mening ensam skulle kunna bli förstådd, ifall den som analysenhet togs ur sitt sammanhang (kontext).

Dessa våra första försök utmynnade i en preliminär regelsamling för

1. segmentering, supplering och strykning av text.

Detta förfaringssätt tränades in och diskuterades, innan vi påbörjade försöken att tolka de enskilda leden inom varje större analysenhet. Utgångspunkten här var Holstis paradigm agent-handling-objekt-mål. Dessa relationer tränade vi oss i att upptäcka och fastställde sedan

2. koder för relationen agent-handling-objekt-mål (oberoende begrepp).

Nya texter gicks igenom då vi försökte tillämpa kodbeteckningarna, och det framstod alltmer tydligt att mycket av vad som dessa texter informerade om bestod i långa begreppskomplex som borde uppdelas för att vi inte skulle få alltför stora enheter att arbeta med. Vi såg oss tvungna att utvidga koderna då Holstis analysmetod visade sig vara otillräcklig för våra syften. Vi skilde ut olika slag av attribut och adverb, och så skapades

3. koder för attribut till agent-objekt-mål samt koder för olika slag av bestämningar till verb (beroende begrepp)

De beroende begreppen relaterades till sina oberoende genom ett tvåsiffrigt kodsystem, där den andra siffran utmärkte underkategorin, beroendet.

Reglerna prövades på nytt från segmentering till tolkning och kodning på olika textavsnitt ur provmaterialet. Denna fas kallar vi

4. en första provkodning.

Den utfördes av två medarbetare i projektet. Efter kodningen gicks steg för steg igenom besvärligheter som uppkommit. Vi resonerade oss fram till en lösning i de olika fallen och sammanfattade denna fas i en exempelsamling med urval ur provintervjuerna. Med ledning av exemplen utarbetades

5. de första kodningsreglerna

som förelåg som arbetsrapport av I. Bierschenk i december 1973 med titeln Innehållsanalys av intervjuer. Dessa regler skulle nu utprövas och så skedde genom

6. provkodning av två av varandra oberoende kodare samt stansning på hålkort.

4.2 De första utprövningsresultaten

I Berg (1974 a) redogörs för överensstämmelsen vid kodningen. Hon använde en lång metod, där varje steg i kodningen från segmenteringen av texten till tilldelning av koder på datablankett bedömdes för sig. Men det prövades också en direkt (kort) metod, där resultatet av den färdiga kodningen bedömdes. Berg (1974 a, s 3) skriver:

> "Den långa metoden är dock ett viktigt steg i inlärningen av kodningsreglerna. Varje steg bör kunna utföras noggrant av kodaren och varje steg bör kunna kontrolleras separat. När väl reglerna är kända kan kodningen ske enligt den korta metoden . ... En sammanslagning eller sammanvägning av reliabiliteten för varje steg bör ge en total reliabilitet för texten. Den korta metoden ger direkt en total reliabilitet."

Resultatet av denna första utprövning skulle vara vägledande för hur en träningsperiod med nya kodare skulle utföras, varför varje steg i kodningsprocessen måste belysas på detta sätt. Sammanfattningsvis gav kontrollen följande resultat.

Segmentering och strykning av text medförde inte några problem. Däremot var suppleringarna olika i ungefär hälften av fallen, beroende på svårigheten i att ge entydiga regler för hur supplering ska ske, dvs vilket begrepp som skall suppleras in som komplement till pronomen, etc. Att supplering har skett på samma ställe av båda kodarna är ett resultat som inte diskuterades i detta sammanhang. Det borde dock ha nämnts, eftersom det visar att reglerna förståtts och tillämpats på samma sätt.

När det gäller tilldelning av koder till begreppen, visade sig de största svårigheterna vid de olika attributen, dvs beroende begrepp. Kriteriet på överensstämmelse sattes till 80 % och i dessa fall uppnåddes endast en 60 %-ig överensstämmelse. De båda metoderna (kort resp lång) visade inga skillnader.

Den första utprövningen gav indikationer på var reglerna borde förbättras, och efter några justeringar utarbetades den första manualen (I. Bierschenk, 1974 a). Nästa steg i utvecklingen var att inskola nya kodare, som inte hade inblick i det arbete som hade föregått manualen. Det textmaterial som projektet förfogar över är av sådan omfattning (ca 4000 sidor) att det ansågs vara av värde att flera personer kunde bearbeta det. Men innan det togs ställning till problemet med flera kodare, bestämdes att en inskolningsperiod med utprövning skulle starta för att ge ytterligare erfarenheter kring dels (1) på vilket sätt reglerna kunde förstås och tillämpas av andra inte involverade personer, dels (2) praktiska problem, såsom miljö- eller tidsfaktorer. I kapitel 5 lämnas en detaljerad redogörelse för denna utvecklingsperiod.

5. KODARTRÄNING OCH PROVKODNING

5.1 Val av kodare

De personer som skulle koda intervjumaterialet måste väljas ut enligt vissa kriterier. De skulle behärska svensk grammatik och syntax och helst i sitt dagliga arbete ha kontakt med språkfrågor och över huvud taget vara sensitiva för språk. En förutsättning var också intresse och vilja att sätta sig in i ett omfattande regelmaterial av det slag som föreligger till denna analys.

Då materialet till övervägande delen berör pedagogiskt forskningsarbete, ansågs det vara en fördel om kodarna hade pedagogisk erfarenhet och inte stod främmande för pedagogisk-psykologiska resonemang.

En viktig faktor i detta sammanhang är också den geografiska anknytningen. Täta kontakter mellan kodarna och forskningspersonalen skulle bli nödvändiga under arbetets gång, varför kodarnas bostad eller arbetsplats ansågs böra ligga inom bekväm radie från forskningsinstitutionen.

Enligt dessa kriterier valdes två personer ut. De hade sitt dagliga arbete som lärare vid Munkhätteskolan i Malmö. Två personer hade redan inskolats i materialet och varit med vid utvecklingen av regelsystemet.

5.2 Instudering och test

I mitten på mars 1974 träffades den grupp som skulle handha kodningen, dvs instruktör, en skolad och två oskolade kodare samt projektledaren för introduktion. De två nya kodarna fick en redogörelse för projektets syfte och materialets utseende och omfång. Vidare redogjordes för principerna för den språkliga analysen. Material av vikt utdelades, såsom frågeformulär, praktiska anvisningar samt Manual för innehållsanalytisk kodning av intervjutexter (I. Bierschenk, 1974 a).

Lektioner med instudering av manualen hölls under veckorna 12-14, sammanlagt 10 timmar, där varje lektionspass om 2 timmar var förberett med textprov, som rättades och diskuterades. Lektionspassen var uppdelade i tre huvudsteg:

1. Segmenterings-, supplerings- och strykningsproblem med följande delmål:
   a. avgränsa relevant textmaterial från icke-relevant
   b. segmentera texten i meningsfullda enheter
   c. identifiera syntaktiska relationer
2. Praktiska åtgärder vid nedskrivning på datablankett samt tolkning av innehåll med följande delmål:
   d. skriva ordenheter i sekvens enligt fastställt mönster
   e. särskilja satstyper enligt definition i manualen
   f. tolka analysenheterna
3. Tillordning av koder enligt innehållsdefinition.

Efter dessa tre steg följde två integrerande lektionspass med nytt textprov som skulle förberedas enligt den ovan givna sekvensen och diskuteras. En första instuderingsperiod betraktades därmed som avslutad.

Två av kodarna hade följt utvecklingen av kodningsreglerna och genomgått ett test och ansågs därför inte behöva testas vid detta tillfälle. De två nytillkomna fick efter de 10 timmarna ett nytt textprov att förbereda (segmentering, uppskrivning på blankett) och koda. Detta skedde i resp kodares bostad med hjälp av manualen. Det bestämdes att provet skulle vara inlämnat i vecka 15.

Testet omfattade en bestämd ip:s svar på en bestämd fråga, sammanlagt 110 ord i 8 grafiska meningar (enligt utskrift).

5.3 Överensstämmelse

Överensstämmelsen hos de två kodarna skulle gälla kategorierna (1) huvudord, totalt, (2) huvudord + attribut hos subjekt/agent, (3) huvudord + attribut hos objekt, (4) enbart attribut, (5) adverb, (6) kod 99/++("innehållslös") samt (7) huvudsats-bisats-relationer. (Kod 88 hade ännu inte tillkommit.) Kategori 6 beräknades inte p g a för liten förekomst. Enbart kort metod användes (se kap 4). Överensstämmelsen i de övriga kategorierna har beräknats med två mått. Det första anger Osgood's index för överensstämmelse mellan bedömare, som beräknas med följande formel (Osgood et al, 1956, s 57):

$$\frac{2 * (\text{bed } 1 + 2 \text{ gem})}{(\text{bed } 1 \text{ tot}) + (\text{bed } 2 \text{ tot})} * 100\,\%$$

dvs relationen mellan det fördubblade antalet gemensamma bedömningar för bedömare 1 och 2, och det totala antalet bedömningar som bedömare 1 och bedömare 2 har gjort.

Med det andra måttet testas med binomialfördelningen huruvida ett i förväg uppställt överensstämmelsekrav på t ex .80 (80%) uppfylls eller inte. Beräkningen utförs enligt följande formel (Siegel, 1956, s 40):

$$z = \frac{x - NP}{\sqrt{NPQ}}, \text{ när } NPQ \geq 9$$

där x = antalet gemensamma bedömningar, N = antalet separata enheter som har bedömts, dvs antalet gemensamma bedömningar + antalet olika bedömningar som bedömare 1 resp bedömare 2 har gjort, P = det uppställda kravet på överensstämmelse, t ex .80 och Q = 1-P = .20. z-värdet testas på

5%-nivån. Observerade z-värden som understiger -1,64 anger att interbedömaröverensstämmelsen inte uppfyller kriteriet. I de fall där NPQ < 9 har Tables of the cumulative binomial probability (1955) använts. Resultatet av kontrollen framgår ur tabell 1.

Tabell 1. Sammanfattning av interbedömaröverensstämmelse vid preparering av text för datorbaserad analys

Innehåll	N	%	z-värden
Huvudord, total	52	100	4.0
Huvudord + attr. (subjekt)	20	91.9	+ .22
Huvudord + attr. (objekt)	15	100	+ 2,70
Attribut	6	66.7	- 2,12
Adverb	17	86.7	- .70
Huvudsats-bisats	17	90.3	- .12

Kontrollen av överensstämmelsen vid prepareringen av text för en datorbaserad analys efter 10 timmars träning visar att tillfredsställande hög överensstämmelse kan uppnås inom samtliga kategorier utom för kategorin attribut sammanslagna. En analys av feltyperna visar att problem uppstår när det gäller att skilja de olika attributen från varandra. Att klassificera huvudord tillsammans med attribut innebär inga svårigheter, men eftersom ett attribut kan stå både före och efter sitt huvudord är kodarna inte helt säkra på kategoritillhörighet (se del III, ss 55-56). Framhävande attribut har förväxlats med satsadverbial, och därav följer att kategorin adverb inte uppvisar lika säkra bedömningssiffror som övriga. Att adverb trots detta visar tillfredsställande överensstämmelse torde bero på att de s k tyngre adverbialleden (rum, tid, sätt och grad) är lätta att identifiera. Ett förhållande som är värt att påpeka gäller kodningen av kategorin huvudord + attribut inom subjekt/agentledet. Att denna kategori inte uppvisar lika hög överensstämmelse som motsvarande objektled kan synas märkligt, eftersom de båda leden i sin syntaktiska struktur (som relationskomplex) är lika. Analysen visar emellertid att det är predikativen som här har vållat problem. Kodarna har vid några tillfällen missbedömt subjektskod vid att-sats efter predikativt verb, dvs att-sats har kodats som objekt efter 'är' (se del III, s 57).

Då testet hade liten omfattning, bör man kanske inte dra för långtgående slutsatser av resultatet. Testet har dock visat att de båda kodarna gjort likadana tolkningar. De felkodningar som gjorts är också lika. Kodarna har således på detta stadium likartad referens till materialet och reglerna.

5. 4 Jämförelse med tidigare överensstämmelsekontroll

För att få en uppfattning om regelsystemets tillämpbarhet för det fortsatta arbetet jämfördes resultatet från denna senare kodarträning med den första utprövningen från januari 1974 (Berg, 1974 a). Samma mått som redovisats ovan användes. Jämförelsen presenteras i tabell 2.

Tabell 2. Jämförelse mellan interbedömaröverensstämmelse vid preparering av text för datorbaserad analys

Innehåll	Jan 1974, kodare (A, B) N	%	z-värden	Apr 1974, kodare(C, D) N	%	z-värden
Huvudord, total	123	89.7	+ .36	52	100	> 4.0
Huvudord + attr. (subjekt)	48	84.3	- 1.38	20	91.9	+ .22
Huvudord + attr. (objekt)	59	77.1	- 3.16	15	100	+ 2.70
Attribut	29	61.9	<- 4.0	6	66.7	- 2.12
Adverb	32	65.3	- 3.64	17	86.7	- .70
Huvudsats-bisats	40	93.3	+ .99	17	90.3	- .12

Berg säger i sin sammanfattning (Berg, 1974 a) att "reliabiliteten för attribut och adverb är dålig" och att "En analys av felen vid kodning av subjekt och objekt med attribut visar att skillnaderna främst beror på fel i kategorierna 30, 33 och 53", dvs subjekt/agent och objekt med efterställt attribut (se del III, ss 55-56). Som tabell 2 visar har såväl huvudord + attribut (objekt) som adverb kunnat nå tillfredsställande överensstämmelse. Klassificeringen av de olika slags attributen kvarstår fortfarande som det största problemet.

Någon exakt jämförelse mellan de två olika tillfällena bör inte göras, dels därför att bedömarna från tillfälle till tillfälle inte är desamma, dels därför att materialet till omfånget varit olika.

5. 5 Provkodning och resultat

Efter denna första instuderingsperiod rådde en viss osäkerhet om fortsättningen av inskolningen. Det beslöts emellertid att kodarna skulle tränas vidare men inom ramen för huvudmaterialet. Instruktören skulle stå till deras förfogande, och de skulle ständigt lämna in material för rättning och diskussion. Samarbete mellan kodarna skulle inte förekomma. Under denna period visade det sig att de båda kodarna arbetade mycket olika, och kontrollerna av kodningsfelen visade större skillnader mellan dem än vad testet gav (I. Bierschenk, 1974 c, kap 5). En av avsikterna med denna period var att få en uppfattning om den tid som krävdes för kodningen. Kodarna upplevde brist på tid och koncentration och det beslöts att perioden skulle avslutas, inte minst av finansiella skäl.

Som en sammanfattning av svårigheterna i kodningsreglerna upprättades ett protokoll på svårighetsgraden (I. Bierschenk, 1974 c, s 7), där det framgår att de största problemen uppstått inom attribut och adverb, liksom under de båda första testen av reglerna. Vidare var suppleringar av så pass differentierat slag att regler för dessa ansågs behöva diskuteras på nytt.

5. 6 Sammanfattande erfarenheter

Förutom vad testet och provkodningsperioden gav för den fortsatta utvecklingen av regelsystemet har kodarna fått ge sina synpunkter på manualen och på det praktiska arbetet. Kodarnas kommentarer finns i I. Bierschenk (1974 c). Perioden i sin helhet kan därför sammanfattas på följande sätt.

För att en kodningsprocedur av det här slaget skall bli effektiv krävs i första hand kontinuitet, även efter instuderingsperioden. Mera styrd provkodning och tätare kontroller än vad som gjorts är nödvändiga för att hanteringen av materialet skall kunna ske med ökad säkerhet. Kodarna behöver särskilt fastställd tid för enbart kodning. Då de båda kodarna som träningen gällde har ett arbete som upptar stor tid dagligen och även ibland under veckosluten, blir kodningsarbetet för splittrat. Regler och förfaringssätt som inlärts tillfredsställande i början kan glömmas bort, om för lång tid förlöper mellan kodningstillfällena. Möjligen torde det vara behövligt att från början systematiskt gå igenom materialet vad gäller problemställningar som behandlas i intervjun, förväntade svarstyper på olika frågor och typ av talspråkstext innan arbetet börjar. Det tycks som om svårigheter kan uppstå på grund av att hela texten (en hel intervju) inte är genomläst, innan kodningen börjar. Detta är viktigt för förståelsen av varje intervjupersons uttryckssätt och för en orientering inom ämnesområdet. Det största praktiska problemet får dock anses vara kodarnas bristande tid. Kodning i avskildhet på institutionen skulle underlätta koncentrationen.

Praktiska åtgärder som underlättar en kodningsprocedur är tämligen enkla att åstadkomma. Ett tillförlitligt resultat beror dock inte i första hand på planeringen av kodningsarbetet eller miljön. Av större vikt är den enskilde kodarens utgångsläge.

Den huvudsakliga skillnaden mellan de båda kodarna gäller förkunskaperna. Bristande språkkunskaper och osäkerhet medför behov av fler "frotteringar" med materialet i starten och att man vid tillämpningen av regelsystemet gärna stöder sig på exempelsamling i stället för att lokalisera kodtillhörigheten med hjälp av rubriker och paragrafer. Även om kodarna från början kunde anses vara "lika", borde förkunskaperna i grammatik och språkstruktur ha testats. Att testet efter den första instuderings-

perioden visar mycket hög överensstämmelse kan ju bero på att man under ett första skede i inlärningen av något nytt får likartad referensram, eftersom man utsätts för samma påverkan samtidigt. Ett initialtest där de båda kodarna hade fått redovisa sina kunskaper utan påverkan av instruktören eller varandra hade troligen visat kunskapsskillnader, som antingen lett till nytt val av kodare eller särskild träning av en av dem före instuderingsfasen.

Till sist kan det vara värt att påpeka att kraven på kodare angående de språkliga kunskaperna borde preciseras. De kunskaper i svenska som en lärare på grundskolans mellanstadium i allmänhet har är inte alltid tillräckliga för detta slag av kodning. Det är antagligen lättare att med goda språkkunskaper som bas kunna sätta sig in i det fackmässiga innehållet i materialet än tvärtom.

6. UTPRÖVNING AV MANUALEN

Efter genomförandet av provkodningarna och på grundval av resultatet justerades manualen från mars 1974 på några punkter. Vi ansåg nu att vi borde utföra en allvarlig utprövning av den. Under den ovan beskrivna träningsperioden visade det sig att de nya kodarna hade behövt en bättre språklig skolning. De två kodarna som varit med om regelutvecklingen har olika språklig bakgrund men ansågs vid denna tidpunkt vara så samtränade, att utprövning av manualen och interbedömaröverensstämmelsekontroll skulle ske med hjälp av dessa båda kodare. Den första kodningen av huvudmaterialet utfördes därför under maj och juni 1974 och omfattade ett frågeområde (ca 10 % av materialet). I Berg (1974 b) redogörs för tillvägagångssättet vid kodningen, principerna för beräkning av överensstämmelse samt resultatet, som ges i tabellform.

Avsikten med denna "reliabilitetskontroll" har varit att testa manualens tillförlitlighet, dvs om kodarna tillämpar reglerna på samma sätt. Kriteriet för överensstämmelse sattes vid 80 %. Ett problem vid bedömningen av beräkningsresultaten är att materialets beskaffenhet (talspråkstext) och vårt krav på varje menings fullständighet medför att vi bör betrakta resultatet från två sidor. Texten prepareras först genom strykning, segmentering och supplering av meningsfragment. Därefter sker tilldelning av koder enligt den avgränsning som kodarna ansett vara meningsfull. Urskiljning av meningar och satser i huvudsats och bisats har inte medfört problem. Men om den ena kodaren har företagit en strykning av ett fragmentariskt yttrande, som den andra kodaren har kunnat tolka och supplera till fullständig mening enligt våra regler, medför detta avvikelser i form av olika antal begrepp, vilket i sin tur kedjereagerar, så att kodarna har olika antal kodade begrepp. Att överensstämmelsen, räknad på båda kodarnas antal begrepp, i detta fall blir otillfredsställande, är helt naturligt. Men eftersom avsikten med kontrollen främst varit att pröva, om regelsystemet kan inläras av två kodare och tillämpas lika, bör man för detta ändamål endast jämföra de båda kodarnas gemensamma text. Denna jämförelse hade varit den enda möjliga om det hade rört sig om skriven text, där struktureringen redan är gjord. Beräknad så, är överensstämmelsen över 80 %, sedd över både intervjuperson och fråga i urvalet. Detta gäller såväl avgränsning av meningar, satser och begrepp som tilldelning av innehållskoder.

Det är framförallt supplerings- och strykningsförfarandet som skiljer vår analys från andra liknande. Därför har det varit viktigt för oss att få belyst vilka svårigheter som finns här. I manualen (del III, kap 4) beskrivs

hur suppleringar sker i de vanligaste fallen. Vid genomgång av de kodade databladen har det visat sig att reglerna följts och att supplering således har skett. De fall som orsakar avvikelser som ger negativt utslag i beräkningen kan belysas med exemplet i ruta 1.

Ruta 1. Olika slag av supplering vid två bedömares tillämpning av kodningsregler

Supplering 1: Det (personkontakter) tycker jag är värdefullt.

Supplering 2: Det (att ha personkontakter) tycker jag är värdefullt.

I det första fallet innebär suppleringen av pronominet 'det' en ersättning genom ett huvudord med bestämning, som enligt kodningen får subjekt/agent-kod. I det andra fallet sätts samma kod på att-frasen och begreppet 'personkontakter' blir objekt. Vad har olikheterna för konsekvenser för analysen? Ja, för det första är suppleringar markerade med parentestecken, vilket garanterar att vi inte analyserar dem såsom yttrade av intervjupersonen. Därför har antalet begrepp inom parentesen ingen betydelse. De båda meningarna, för det andra, säger att en person (ip) tycker att någonting är värdefullt, nämligen personkontakter. Analysen har ännu inte kommit så långt att det finns en regel för hur kopplingen mellan 'värdefullt' och 'personkontakter' ska ske i de bägge fallen. En alternativregel måste skapas, som tar hand om huvudbegreppet i huvudsats respektive samma begrepp i bisats, när ett predikativ föreligger. För att åskådliggöra en möjlig koppling kodas dessa båda meningar i ruta 2.

Ruta 2. Kodning av mening vid två bedömares avvikande supplering

Supplering 1	kod	Supplering 2	kod	
Det (personkontakter)	30	Det (att)	30 2	
tycker jag	42	(ha)		40
är	41	(personkontakter)		50 1
värdefullt	32	tycker jag	42	
		är	41	
		värdefullt	32	

För kodernas innebörd hänvisas till del III, kap 5.

Det oberoende begreppet (huvudbegreppet) är i bägge fallen 'personkontakter'. I det första fallet står det med subjektskod i huvudsats, i det andra fallet som objekt i bisats, av den anledningen att 'att' endast indikerar ny sats och inte självt uppbär 'innehåll'. Om vi vill ha reda på vad som är värdefullt, kan vi med en sökning koppla predikativet (32) med närmast befintliga huvudbegrepp, dvs som har huvudkod 3, i samma sats eller vid

'att' efter verbet i bisats (det kunde ju ha funnits ett subjekt före verbet). Sådana möjliga kopplingsregler kan dock först skapas, efter det att materialet finns samlat. Vad jag här har velat visa är endast att olikheter vid suppleringar kan komma att upphävas.

Andra olikheter kan gälla ett personligt pronomen (han, hon, etc) som supplerats med t ex 'forskare' respektive 'assistent'. Analysen kommer att ta hänsyn till kategorier av begrepp, t ex kategorin forskarpersonal, dvs personer som arbetar med forskning. Supplerade begrepp som tilldelas samma kategoribeteckning kommer då att betraktas som likvärdiga. Sådana avvikande suppleringar har i vår kontroll markerats som olikhet, vilket har bidragit till sämre överensstämmelse. Kategoriseringen är ännu inte påbörjad, varför vi inte har kunnat kontrollera suppleringarna ur denna aspekt.

Generellt om suppleringsregler skulle kunna sägas att de i högre grad än nu skulle kunna formaliseras, t ex borde det inte vara ogörligt att ge en regel för om och när att-fras eller -sats skall användas. Exemplet i ruta 2 torde klart visa att alternativ 2 är att föredra, ifall man låter kongruensböjningen (värdefullt) fälla avgörandet. Vi har försökt att skära ner antalet långa suppleringar genom standardiseringar av intervjufrågorna (se del III, s 52). Berg (1974 b, s 31, preliminärversion) skriver:

"Vid en kommande revidering av manualen bör suppleringsförfarandet i ännu högre grad standardiseras."

Jag är inte säker på att jag delar den uppfattningen. Standardisering i det här fallet innebär att en att-sats (som referens till ett 'det', det vanligaste slaget av avvikelse) komprimeras till oberoende begrepp plus ett beroende i form av efterställt attribut (modifierare). På detta sätt får vi ett verblöst komplex. Eftersom vi är intresserade av verben som bärare av handlingar eller tillstånd, bör de, särskilt när vi ändå måste supplera, finnas med, för att vi skall få veta vilka de är. Först efter det att satser bildats, kan vi komprimera eller standardisera om vi finner det lämpligt för uppbyggandet av register. De som redan gjorts är försvarbara, eftersom de avser de frågor som intervjuaren ställer till varje ip och som har känd referens genom hela materialet.

För att angripa problemet med suppleringar av den typ som diskuterats ovan bör man undersöka i vad mån olika sätt att supplera samma innehåll innebär svårigheter vid en maskinell sökning. Den syntaktiska positionen och tilldelad innehållskod bör samverka vid denna kontroll.

Kontrollen på överensstämmelse mellan de två kodarna har visat att manualen i sitt nuvarande skick har fungerat tillfredsställande. Reglerna har använts lika av båda kodarna. Det påpekades under kapitel 5.6 att

goda språkkunskaper var nödvändiga. Trots att kodarna ansåg sig vara olika språkligt kapabla, har överensstämmelse ändå uppnåtts, tack vare perioder av samträning, där intresset har haft stor betydelse.

En analys av resultatet för de olika begreppen visar att det i några fall ändå förekommer tveksamma kodningar. Berg (1974 b, bilaga 2:16) redovisar värdena för varje enskild innehållskategori. Liksom tidigare bereder adverben störst svårigheter. Det är den enda kategorin, där överensstämmelsen är lägre än 80 %. Man kan emellertid utläsa ur tabellen att resultatet blir bättre ju längre kodningen framskrider. Den sist kodade intervjun, som har flest antal adverbkoder, visar tillfredsställande överensstämmelse.

Interbedömaröverensstämmelse eller likhet i bedömning mellan flera bedömare är ett viktigt resultat i en metodutveckling av det här slaget. Men en sådan kontroll säger ingenting om det kodade och stansade materialets korrekthet. Granskning av materialets skick har skett som det sista ledet i denna första fas av utvecklingen. Dels har en kontroll av säkerheten i stansningen genomförts, dels har koderna i datautskrifterna granskats mot kodningsreglerna och rättats, för att vi i fortsättningen ska ha ett perfekt material att basera våra första analyser på.

7. KONTROLL AV STANSNINGS- OCH KODNINGSFEL

Då överensstämmelsekontroll av det kodade avsnittet visat tillfredsställande resultat, har vi ansett oss kunna övergå till nästa steg i utvecklingen av analysen, nämligen att kontrollera materialets skick, sådant det nu föreligger. Det är i detta skick det i fortsättningen skall hanteras, och därför måste det vara felfritt.

Detta kapitel avser att granska materialet ur två aspekter. Dels kan det stansade materialet innehålla fel, trots att en korrekt kodning har skett. Dels kan stansresultatet innehålla fel som beror på inkorrekt kodning, således ej på stansoperatrisen. För att få veta i vilket skick materialet befinner sig har genomförts en kontrollstansning av en del av det stansade materialet och en granskning av i vilken utsträckning de kodade och stansade enheterna överensstämmer med kodningsreglerna i hela materialet. Syftet med granskningarna kan formuleras på följande sätt:

1. Syftet med kontrollstansningen är att få ett felfritt stansat material samt att
   - 1.1 jämföra två olika stansningar med varandra för att avgöra pålitligheten i stansningen
   - 1.2 få överblick över vilka de eventuella stansfelen är och i vilken grad de förekommer
   - 1.3 försöka förklara orsakerna till förekomsten av stansfel för att kunna ge rekommendationer för det fortsatta arbetet
2. Syftet med kontrollen av kodningsfel är att
   - 2.1 granska att konsekvens i kodningen upprätthålls genom hela materialet
   - 2.2 jämföra fel och ändringar i materialet dels inbördes, dels med avseende på stansningsfel och kodningsfel.

7.1 Kontrollstansning

7.1.1 Urval och förfaringssätt

Det material som för närvarande ligger på magnetband vid Lunds Datacentral (LDC) är intervjumaterialets (40 intervjuer) kodade frågor 5-8 (ca 10 % av det totala materialet) omfattande 36 689 hålkort. Stansningen skedde av stansoperatrisen vid lärarhögskolan i Malmö i två perioder under juni och juli 1974. Hålkorten överlämnades sedan till Lunds Datacentral tillsammans med instruktioner för kontrollstansningen. Sådana var nödvändiga, framför allt därför att kodarna utvecklat rutiner för en "ekonomisk" kodning, som gäller vissa kolumner. Rutinerna framgår ur manualen (del III, ss 66-76).

Det totala antalet datablad i detta material är 1450. Samtliga är dock inte fullskrivna, beroende på strykningar och korrigeringar under kodningens gång. I genomsnitt fanns det 30 blad per intervjuperson (ip). Kontrollstansning skulle ske på 10 % av det av LHM stansade materialet, alltså 3 blad per ip. Med hjälp av slumptabell bestämdes att det första bladet hos varje ip skulle vara nummer 5. Varje ip:s totala antal blad delades med siffran 3, varpå intervallen mellan de tre bladen beräknades på så sätt som framgår ur ruta 3.

Ruta 3. Exempel på urval av material för kontrollstansning

ip nr	antal blad/ip	$\frac{\text{antal}}{3}$	kontr. stans blad nr
1	37	12	5, 12, 29
40	12	4	5, 9, 1

De 120 bladen sorterades ut för att materialet skulle bli mera lätthanterligt för den nya stansningen. Kontrollstansningen genomfördes på LDC under september månad 1974.

Efter kontrollstansningen jämfördes denna med det vid LHM stansade materialet. Denna jämförelse utfördes maskinellt. Resultatutskrifterna innehöll en listning av kort med avvikelser. Listningen redovisar främst två typer, dels (1) de kort, där identifikationskoderna avviker (kolumn 1-10), dels (2) de kort, där övriga koder (tema, text och innehållskoder, kolumn 11-80) avviker. Skillnaden mellan båda typerna är att den förra inte möjliggör jämförelse inom det stansade kortets återstående kolumner, då det inte har kunnat återfinnas på sin rätta plats. Utskriften redovisar endast LDC stansning, men det innebär inte att kortet saknas i LHM stansning. Det har p g a felstansning i de första kolumnerna sorterats in på fel plats. Ytterligare en maskinell kontroll gjordes därför, som avsåg följden i materialet, dvs att ip-nummer, frågenummer och meningsnummer följer i ordning efter varandra och att de ingående orden har rätt ordningstal. Först efter registrering och rättning av denna feltyp, kunde kort med avvikelser inom typ 1 jämföras till fullo. Inom avvikelsetyp 2 har samtliga nedslag på hela kortet direkt kunnat jämföras. Nedan ges exempel på de två avvikelsetyperna.

Ruta 4. Exempel på avvikelser enligt utskrift efter kontrollstansning

Avvikelse-typ 1	KONTROLLSTANSAT KORT HAR INGEN MOTSVARIGHET I HUVUDMATERIALET:			
	03050016121 21	EN MÄNGD	51	
Avvikelse-typ 2 a	LDC: 04050011061 11	TUESENTOALS	51	
	LHM: 04050011061 11	TUSENTALS	51	
b	LDC: 04080011081 21	STÅR	40	
	LHM: 04080011081 21	STÅR		40
c	LDC: 050800220211 11	ÄR	41	
	LHM: 05080022021 11	ÄR	41	

Felen markerade med ring.

Registreringen och markeringen av de i ruta 4 exemplifierade avvikelserna redovisas längre fram (Ruta 7).

7. 1. 2 Feltyper: kategoriindelning och registrering

De typer av fel som förekommer har indelats i fyra huvudkategorier, A, B, C, D med underkategorier (typer), som framgår ur nedanstående uppställning.

Ruta 5. Översikt över feltyper och deras innebörd vid beräkning av kontrollstansningsresultat

Kat.	Typ	Innebörd	Kat.	Typ	Innebörd
	A 1	ip nr		B 1	källa
Identi-fikation	2	fråga nr	Tema	2	negation
	3	mening nr		3	tempus
	4	ord nr		4	modus
				5	övr. satstema
	C 1	stavning			
Text	2	parentes	Inneh.	D 1	begrepp
	3	annan text		2	satskolumn

Fel av typ A1-4 gäller identifikation. 1, 2 och 4 tar två kolumner i anspråk och 3 omfattar fyra, sammanlagt kolumnerna 1-10 på datablanketten. (Kolumnerna 24-25 räknas också hit. Dessa används endast när ett begrepp omfattar två kort, således relativt sällan, se manualen kap 6, s 65).

Typ B 1-5 avser kolumnerna 11-23, tema koder. B 1 anger vem som står för yttrandet (talaren själv eller refererande till annan källa). B 5 anger satsteman, såsom fråga, villkor, etc.

C 1-3 gäller text, dvs den icke-numeriska delen, som omfattar kolumnerna 26-65. Avvikelser inom denna kategori resulterar främst i C 1, stavning, som avser feltolkningar och felslag av bokstav, omkastning av bok-

stäver, extra bokstav, saknad bokstav, mellanrum mellan bokstäver i ord respektive ihopskrivning av ord, samt felvänd parentes. C 2 betecknar glömd parentes eller stansad parentes som inte är kodad. Parenteser är talrika i materialet, beroende på suppleringar. Detta berättigar att parenteserna bildar en egen kategori. C 3 innebär att ett ord inom ett begrepp saknas eller att ett ord för mycket har stansats. En feltolkning som innebär att ett helt annat ord har stansats (t ex oerhört - säkert) faller också inom denna typ. Likaså räknas hit även ändelser som innebär ett annat ord, som skulle få annan innehållskod (t ex forskar - forskare). Undantag gäller dock vid typen kan - han, har - hav, där det har utgåtts ifrån att de förväxlade bokstäverna orsakats av otydlig kodning. Slutligen faller inom C 3 avvikelser i form av "språkbyte", t ex att ett utländskt begrepp har tolkats som ett svenskt (summeries - summering).

D 1-2 markerar fel inom innehållskodningen i kolumnerna 66-80. D 1 avser kodbeteckning plus hänvisningskod (loop) och D 2 kodens placering i satskolumn ("spalt"), dvs huvudsats-bisatsplacering, där varje satskolumn innehåller två siffror och loop-koden en.

Efter bestämning av kategorierna gicks utskriften igenom, och varje avvikelse registrerades med respektive beteckning. Men för att ett protokoll med frekvenser skulle kunna upprättas, måste det först bestämmas på vilket sätt feltyperna skulle bedömas. Nedanstående ruta visar principerna för bedömning av fel inom respektive feltyp.

Ruta 6. Feltyper och deras registrering vid beräkning av kontrollstansningsresultat

Kat	Feltyp	Registrering	Markering
Identifikation	A 1-4	Reg. 1 gång per fall Systematiskt fel reg. 1 gång (Skilln. som uppstår genom en- resp två-siffriga tal reg. inte.)	1
Tema	B 1-5	Reg. 1 gång per fall Systematiskt fel reg. 1 gång	1
Text	C 1-2	Reg. 1 gång per fel/saknat nedslag	1
	C 3	Reg. 1 gång per fall samt 1 fel per fel /saknat/ extra nedslag	1 (1-n)
Innehåll	D 1	Reg. 1 gång per fel /saknat/ nedslag	1
	D 2	Reg. 1 gång per fall samt 1 fel per fel /saknat nedslag	1
		Samtliga kolumner som felet omfattar (minst 4, högst 6)	1 (4, 5, 6)

De exempel från den maskinella utskriften som gavs i Ruta 4 skulle följaktligen registreras på följande sätt.

Ruta 7. Exempel på registrering och markering av fel enligt avvikelser vid kontrollstansning

		Registr.	Mark.
Fel av typ 1	LHM:	A 4	1
Fel av typ 2 a	LDC:	C 1	2
2 b	LDC:	D 2	1 (4)
2 c	LHM:	B 2	1

Det fel som registrerats för LDC avseende D 2, placering i satskolumn, markeras som 1 fel och anger samtidigt över hur många kolumner felet sträcker sig. Koderna är tvåsiffriga, och två siffror har stansats som extra kod i en satskolumn och saknas i en annan satskolumn, dvs 4 kolumner är felkodade. Det bör påpekas att fel per kolumn inte räknas inom typerna C 1, stavning, och C 3, annan text, i fel vid extra eller saknade nedslag. I fall som:

NÅGON S ¦ DÅ DÄR
NÅGON S ¦ Å DÄR

tar den första stansningen en kolumn mer i anspråk och nedslagen från den streckade linjen skulle samtliga avvika, räknat per kolumn. Inom textavsnittet är vi inte intresserade av de enskilda kolumnerna utan endast av den information som ges inom varje begrepp, dvs den text som står på varje kort. Kolumnerna har en viktigare funktion inom D-kategorin, där varje sats löper i "slingor" och där en felplacering i satskolumn får konsekvenser vid den maskinella läsningen av meningar och satser.

7.1.3 Resultat

Registreringen och markeringen av de avvikelser som förekommit i LHM och LDC stansningar har givit de resultat som presenteras i tabell 3. Men för att en diskussion skall kunna föras kring de observerade felen, beräknades också den relativa frekvensen. Det totala antalet kontrollstansade blad är 120. Av dessa uttogs de 12 första bladen, dvs 10 %, för beräkning av det totala antalet nedslag. (Slumpmässigt urval föreligger redan, varför ett sådant förfarande inte var nödvändigt på nytt.) Antal nedslag per 12 blad är 7026. Antalet nedslag per de 120 bladen är således 70260. Andelen fel beräknades dels per kategori (A, B, C, D), dels för kategorierna sammanslagna. Resultatet av dessa beräkningar presenteras i tabell 3.

Tabell 3. Jämförelse mellan LHM stansning och LDC kontrollstansning av 10 % av materialet. Observerad och relativ frekvens (beräknad på antalet nedslag)

Kat.	Typ	LHM f	LHM %	LDC f	LDC %
Identi-	A 1	-	-	-	-
fikation	2	-	-	-	-
	3	2	.00	1	.00
	4	5	.01	3	.00
Tema	B 1	-	-	1	.00
	2	1	.00	1	.00
	3	-	-	2	.00
	4	1	.00	1	.00
	5	1	.00	3	.00
Text	C 1	20	.03	75	.11
	2	6	.01	8	.01
	3	-	-	19	.03
Innehåll	D 1	-	-	3	.00
	2	22	.03	29	.04
Tot.		58	.08	146	1.21

Ur tabell 3 framgår en tydlig skillnad mellan LHM och LDC stansningar till fördel för LHM. Likheterna är störst inom kategorierna A, identifikation, och B, tema, skillnaderna störst inom kategori C, stavning. A och B gäller numeriska koder, där antalet numeriska tecken som kan förekomma är mycket begränsat. Endast A 3-4 (menings- och ordnummer) varierar i högre grad än de övriga. Där finns också tydligen större möjligheter till felslag. En viss skillnad råder mellan A och B. A 1-2 gäller siffror, som systematiskt skall stansas, ibland över flera blad. Där finns inga felslag. A 3, dvs meningsnummer, gäller också systematisk stansning, men inte i lika hög grad som i A 1-2. A 4 avser ordnummer i ordningsföljd och där uppstår lättare ett fel. Systematiken i tema-kategorin är av ett annat slag. B 1, 3 och 4, dvs källa, tempus och modus, stansas alltid, B 2 och 5 (negation resp övrigt satstema) inte alltid. Dessutom spänner B 5 över många kolumner, och felslag här torde kunna vara lättare (t ex att nedslag görs i fel kolumn) än inom de övriga B-typerna. Även om skillnaderna är ytterligt små mellan LHM och LDC när det gäller dessa två kategorier, kan det dock sägas att LHM stansning är pålitligare, framför allt om man betänker innebörden i kategorierna och konsekvenserna av felslag. B-kategorin (tema) måste betraktas som viktigare än A (identifikation), eftersom temat uttrycker tolkning av texten.

Fel inom A kan lätt identifieras genom en maskinell kontroll. För att kontrollera typ B fel måste man undersöka texten, vilket är oerhört mera tidsödande och besvärligt. Till detta skulle fordras ett omsorgsfullt utarbetat program.

Kategori D (innehållskodning) tillhör också den numeriska stansningen. Någon systematisk stansning förekommer dock inte, och variationen inom de förekommande tecknen är något större än i de ovan nämnda.

Den största skillnaden mellan A och D är att D inte kan tolkas av stansoperatrisen. Därtill skulle fordras kännedom om regelsystemet. Koderna inom D 1 är väsentliga, eftersom de utgör en klassificering av textens olika begrepp. Att stansa ett fel här, så att t ex ett verb kommer att tillhöra någon kategori substantiv (subjekt, objekt), får allvarliga konsekvenser. Det tar lång tid att gå igenom utskrifter för att kontrollera alla begrepp som ingår i respektive kod. Detta bör man ha i åtanke när man betraktar den till synes lilla skillnaden i D 1 mellan LMH och LDC.

När det gäller D 2 (satskolumn) enbart, är skillnaden mellan LHM och LDC mycket liten. LDC har gjort två felplaceringar mer än LHM (6 (4) resp 4 (4)). Fel här är också allvarliga, då placering i kolumn anger i vilken sats begreppet står. Men genom att lokalisera t ex tre verb till en och samma mening, går det relativt bra att upptäcka fel. Verbkoden är i sig själv viktigare. Men med tanke på kolumnernas stora antal måste man nog räkna med att ett visst antal fel kommer att uppstå här. Jämfört med felberäkningsprinciperna (extra stansade + tomma kolumner, enligt Ruta 6) ser värdena i tabellen högre ut än vad felet möjligen berättigar till.

Det återstår att diskutera kategori C (stavning), som visar de intressantaste skillnaderna mellan de båda stansningarna, C 2 kan betraktas som lika. Parenteser förekommer ofta i texten. Dessa få parentesfel kan inte negligeras. I samtliga fall har den avslutande parentesen antingen glömts eller stansats, trots att parentes framför begreppet saknas. Om vi vill lista samtliga begrepp som står inom parentes, kan vi fortfarande arbeta med materialet, om den första parentesen finns med, men där den saknas kommer hela begreppet att saknas i utfallet. Parentes kring begrepp anger att supplering har skett. Sådana är talrika i materialet och indikerar att begreppet inte har uttalats av intervjupersonen (se manualen kap 4). Parentestecknet är således en viktig signal när vi skall tolka reella utsagor kontra implicerade eller tänkta. En kontroll i de kodade bladen visar att 5 av dessa 14 parentesfel är orsakade av kodarna själva. Det tycks vara lätt att förbise framför allt den avslutande parentesen. En förklaring till varför stansoperatriserna trots kodarnas förbiseende ändå inte har

fullföljt parentesen kan vara att man inte har känt till innebörden i tecknet i det här fallet.

Inom C 3 uppvisar LHM en lika säker stansning som i D 1. Jämförelsen görs därför att även C 3 är en väsentlig kategori. Varje enskilt begrepp (som återfinns på ett hålkort) är identifierat och skall kunna tas ur materialet utan att förlora identiteten, dvs tillhörighet till ip, frågeområde, meningsnummer eller ordnummer. Begreppet tillsammans med sin innehållskod är det komplex vi skall arbeta med för vidare analys. Om texten lyder 'forskare' och har koden '40' säger detta kort oss inte var felet ligger; fel text? eller fel kod? Därför kan fel av typ C 3 bli besvärliga att identifiera. C 3 och i ännu högre grad C 1 visar kontrollens största skillnader. När man arbetar med text är det synnerligen tråkigt om det finns stavfel i den. En rätt stansad innehållskod kan förhindra feltolkningar, men vi kommer inte alltid att arbeta med text tillsammans med kod. Arbete med text i dator kräver korrekt sammansättning av alfabetiska tecken. Även här tycks vi i hög grad kunna lita på LHM.

De numeriska stansningarna per kort är i genomsnitt fler än de med alfabetiska tecken. Trots detta förekommer de flesta felen vid text. Vid förfrågan på såväl LHM som LDC har det sagts att man nästan uteslutande arbetar med numeriskt material, vilket är en trolig förklaring till resultatet. En mera djupgående orsak skulle kunna vara att människor i allmänhet förhåller sig annorlunda till siffror än till bokstäver (som bildar ord). Ord stansas antagligen snabbare, eftersom man läser ut hela ordet innan man stansar det. Risken för fel blir då större. I allmänhet vet man inte vad de enskilda siffrorna står för, vilket innehåll de representerar. En "större respekt" för siffror medför långsammare stansning och därmed också mindre antal fel.

Avslutningsvis kan man betona att LHM, även när det gäller texten, stansar säkrare än LDC. Något kan detta förklaras med att personlig kontakt mellan stansoperatrisen och kodarna har kunnat hållas, så att eventuella oklarheter i texten har kunnat förtydligas. Men dessa har i förekommande fall korrigerats, vilket också kommit LDC till godo. Det kan dock inte förnekas att stansoperatrisen vid LHM på detta sätt blivit förtrogen med materialet och kodarnas handstil, vilket LDC inte haft möjlighet till.

I vad mån handstilen hos kodarna har kunnat orsaka stansfel eller om felen under C är rena felslag kan endast avgöras efter kontroll av varje enskilt fel mot de kodade bladen. En sådan kontroll bör vara av betydelse för det fortsatta arbetet, varför textavsnittet skall granskas närmare med avseende på arten av fel.

7. 1. 4. Granskning av textfel

Vi har konstaterat att fel i kategorierna C (text) och D (innehållskodning) får de allvarligaste konsekvenserna för materialets pålitlighet, och att textfel lättare uppstår än numeriska fel. I det följande skall granskas vilka felslag inom kategori C som kan hänföras till feltolkning p g a kodarnas handstil och vilka som är stansarnas egna felslag. Typ C 2 (parentes) och 3 (annan text) granskas inte, dels därför att C 2 förekommer i förhållandevis liten omfattning och redan har diskuterats, dels därför att C 3 innebär en typ av avvikelse som knappast kan bedömas ur denna synvinkel. Vi granskar således typ C 1 : stavning (innebörden redovisas i Ruta 4).

Det kodade materialet är kontrollerat, och inga stavfel förekommer inom det aktuella urvalet. Feltyperna saknad och extra bokstav, mellanrum i ord, hopskrivning av ord, felvänd parentes och omkastning av bokstäver utelämnas här. Dessa ingår senare i redovisningen som "övriga fel". De granskade feltyperna redovisas som "felslag" respektive "tolkningsfel", som kan bero på otydlig kodning. För att någorlunda säkert kunna hänföra felen till endera typen kontrollerades de först mot stansarnas tangentbord, som har det utseende som figur 1 visar. (Endast tecken av intresse för denna analys redovisas.)

Efter kontroll av felen jämförda mot tecknens placering på tangentbordet kunde flera med stor säkerhet hänföras till "felslag". Andra granskades mot kodningarna på databladen och fel som med stor sannolikhet beror på kodarnas handstil hänfördes till "feltolkning". En tredje grupp återstod, som visade sig vara svår att föra till endera gruppen. De rubriceras som "osäkra". Ruta 8 ger en översikt över indelningen och de observerade felen.

Uppställningen i ruta 8 visar att det totala antalet felslag respektive feltolkningar är ungefär lika. Felslagen är lätt förklarbara vid en jämförelse med figur 1. Antingen förväxlar man tecken som ligger vid sidan om varandra eller så slår man ned den tangent som ligger över eller under tecknet ifråga.

Det torde också vara klart att feltolkning beror på otydlig handstil: för långa eller för korta staplar (t ex h/n), r tolkas som v, osv. I anvisningarna för kodningen sägs att varje bokstav skall skrivas för sig i avsedd kolumn. Men det förefaller naturligt att stilen efter hand "flyter ut" till s k skrivstil, vilken i allmänhet brukar vara svårare att läsa. Vid en kontroll av respektive fel märks tydligt att många av stansfelen beror på att kodarna har skrivit snabbt. (Dåligt vässad penna kan säkert bidra.)

								—	0
+ Q	W	) E	R	T	Y	1 U	2 I	3 O	P
A	S	D	F	G	H	4 J	5 K	6 L	
Z	X	C	V	B	(N	7 M	8 Å	9 Ä	Ö

Figur 1. Tangentbord för hålkortsstans IBM 29. De för analysen relevanta tecknen

Ruta 8. Översikt av textfel vid jämförelse av LHM stansning och LDC kontrollstansning. Kodade och stansade tecken

	Felslag		Feltolkning		Osäkra	
LHM	Kodat	Stansat	Kodat	Stansat	Kodat	Stansat
	på	oå	man	men	linguistics	lingeistics
	i	2	vad	val	vad	vak
	gäller	fäller	att	ett	då	de
	och	3ch	än	an		
	finns	finna				
	som	dom				
	(som)	7 som)				
LDC	Felslag		Feltolkning		Osäkra	
	Kodat	Stansat	Kodat	Stansat	Kodat	Stansat
	som	dom	här	när	man	kan
	fram	frsm	Ulf	Ult	ja	ta
	projektet	projekter	ja	ju 2)	förhållan-det	förhållan-den 4)
	sökte	säkte	har	var	av	om
	undervisn.	indervisn.	bevistat	beristat	review	deview
	är	år 1)	höra	köra	gett	sett
	gången	fången	utav	utan	relativt	reletivt
	håller	hållet	förändrat	förandrat		
	kybernetik	kybermetik	vilja	vi(ja		
	syfte	syfte	Beredsens	Beredcens		
	hand	hsd	hatt	haft		
	tre-betygare	tre-betugare	håller	hållet		
			psychologi-cal	psyckologi-cal 3)		
			kan	han		
			empirisk	empivisk		
			UPS	VPS		
			kände	hände		
			har	hav		

1) förekommer 3 ggr 2) 4 ggr 3) 3 ggr 4) räknas som 2 fel

Gruppen "osäkra" är som nämnts svårare att förklara. Några otydligheter har inte kunnat upptäckas på databladen, och läget på tangentbordet tycks inte heller spela in. Möjligen kan två fall (m/k, r/d) hänföras till felslag, eftersom tecknen ligger nära varandra, i diagonal.

Sammanfattningsvis bör framhållas att texten är sammanhängande och fullt förståelig, läst uppifrån och ner på bladet. Att feltolkningar ändå görs på så sätt att nya begripliga ord bildas (t ex kände/hände), som inte passar in i sammanhanget, kan förefalla märkligt. Men detta kan förklaras dels genom att stansoperatriserna torde vara främmande för en stor del av innehållet i textmassan, dels och kanske främst genom att maximalt varje kort ses som en enhet och stansas för sig utan att man reflekterar över de föregående korten. Att LDC genom urvalet inte alltid haft hela meningar att arbeta med skulle i så fall inte spela någon roll. Det hela rör sig helt enkelt inte om någon tolkningsprocess och skall heller inte göra det.

Till sist skall redovisas beräkningar av antalet fel inom typ C 1 (stavning), där vi jämför feltolkningar med övriga kategorier. Dels redovisas andelen fel inom kategorin (% C 1), dels andelen totalt (% t).

Tabell 4. Jämförelse mellan LHM stansning och LDC kontrollstansning av 10 % av materialet. Feltyper inom C 1 : stavning. Observerad och relativ frekvens (beräknad inom C 1 och totalt)

Kat. C 1	LHM			LDC			Tot. C 1		
	f	%C 1	% t	f	%C 1	% t	f	%C 1	% t
Feltolkning	4	4.2	.01	23	24.2	.03	27	28.4	.04
Felslag	7	7.4	.01	14	14.7	.02	21	22.1	.03
Osäkra	3	3.2	.00	7	7.4	.01	10	10.5	.01
Övriga fel	6	6.3	.01	31	32.6	.04	37	38.9	.05
Tot. C 1	20	21.1	.03	75	78.9	.10	95	100	.13

Feltolkning räknas som kodningsfel
Övriga kategorier räknas som stansfel

Tabell 4 visar att de fel inom textavsnittet som uppstått till största delen beror på stansarna själva. Ca 28 % av felen (eller .04 % av nedslagen) skulle kunna bero på kodningen. Felstansningarna utgör hos LHM ca 17 % (eller .02 % av nedslagen). Om resultatet beror på en i allmänhet större säkerhet hos stansoperatrisen på LHM är svårt att fastslå. Om vana vid materialet kan ha så gynnsamma effekter på så kort tid, har vi skäl att vara nöjda. Vilka orsakerna än kan vara till det bättre resultatet för lärarhögskolans stansoperatris konstaterar vi att hennes stansning är säkrare för materialet än om vi hade lämnat över det till LDC.

7.2 Överensstämmelse i kodning och stansning

7.2.1 Omfattning och registrering

Såsom nämnts visade utskriften över avvikelser efter kontrollstansningen en feltyp som avsåg följden i materialet, dvs fel i identifikationskoderna. Det sades också i diskussionen under tabell 3 att sådana fel kunde kontrolleras maskinellt. Denna kontroll genomfördes vid LDC under oktober månad 1974. De felaktigt insorterade korten korrigerades, men dessa fel har i fortsättningen inte markerats. Den totala kontrollen bygger i stället på kategorierna B (tema), C (text) och D (innehåll) enligt den tidigare redovisade indelningen. Avsikten med denna nya kontroll är främst att granska innehållskodernas konsekvens genom hela materialet. Kodningen har skett under en längre tid, och vissa inkonsekvenser kan ha uppstått p g a ändrade referenser hos kodarna och att vissa koder har visat sig vara svårare att skilja på än andra. Hela materialet har således rättats, varför det också blev möjligt att protokollera felen för att få en uppfattning om i vilken utsträckning de beror på stansfel respektive kodningsfel. Samtliga 40 intervjupersoners hela text, avseende det aktuella frågeområdet, gicks igenom kort för kort, och varje fel registrerades på samma sätt som redovisats i ruta 6, med de undantag som redogörs för i ruta 9.

Ruta 9. Särskilda felregistreringsprinciper gällande vid kontroll av kodnings- och stansfel inom det av LHM stansade materialet

Feltyp	Registrering	Markering
C 3	Reg 1 gg per fall	1
	Ändring i text: ord stryks eller läggs till, 1 gg per ord	1
	Omplacering: ord byter plats p g a kodbyte. 1 gg per ord	1
	När p g a fel av typ C 1 och C 2 ett ord måste omplaceras registreras det inte	-
	När begrepp förs ihop till ett kort och kod sammanfaller registreras det inte (se D 1)	-
D 1	Ändring i underkategorin (t ex 44 ⟶ 46) reg. som 1 fel	1
	Ändring i huvudkategori (t ex 50 ⟶ 70) reg. som 1 fel	1
	Ändring i huvud- och underkategori (t ex 44 ⟶ 53) reg. som 2 fel	2
	Ändring av verbkod som delas genom annan kod reg. 1 gg	1
	Ändring vid begreppskomplex (t ex 71, 70, 73 ⟶ 51, 50, 53) reg. som 1 fel	1
	När två koder sammanfaller till en (t ex 31, 32 ⟶ 32) reg. det som 1 fel	1

Såsom exemplifieringen i ruta 9 visar, beror registreringen under C 3 huvudsakligen på ändringar som har ansetts nödvändiga för att materialet skall vara så enhetligt som möjligt. Justeringarna avser främst suppleringar och ny ordavgränsning, vilket i sin tur även påverkar kodningen i D 1. När det gäller D 1 medför beräkningsprinciperna att denna feltyp kommer att redovisas separat, för att ändringar och fel inom de olika innehållskoderna för begreppen skall kunna åskådliggöras och diskuteras.

7.2.2 Resultat

Granskningen av fel i hela det stansade materialet har givit de resultat som redovisas i tabell 5. Den relativa frekvensen är beräknad på antalet nedslag totalt, dvs 702 600. Stansfelen rubriceras i tabellen LHM och kodningsfelen K.

Tabell 5. Jämförelse mellan stansfel (LHM) och kodningsfel (K) vid granskning av det totala stansade materialet. Observerad och relativ frekvens (beräknad på antal nedslag totalt)

Kat.	Feltyp	LHM f	%	K f	%	Tot f	%
Tema	B 1	-	-	-	-	-	-
	2	3	.00	47	.01	50	.01
	3	-	-	1	.00	1	.00
	4	1	.00	2	.00	3	.00
	5	2	.00	3	.00	5	.00
Text	C 1	173	.02	9	.00	182	.03
	2	9	.00	25	.00	34	.00
Inneh.	D 2*	88	.01	99	.01	187	.03
	Tot	276	.04	186	.03	462	.07

*D2:för markering, se ruta 6
( ) anger antal kolumner, t ex 2(2) = 2 fall som påverkar 2 kolumner

LHM: 2(2), 17(4), 2 (8)
K: 6(2), 1(3), 14(4), 2(5), 3(6)

Som framgår ur tabell 5 beror den större delen av felen på LHM stansning. De flesta korrigeringar har skett inom C 1, D 2 och B 2, dvs stavning, satskolumn och negation. Inom C 1 beror dessa nästan uteslutande på stansfel.

Korrigeringarna inom D 2 samverkar med ändringar i texten (K = 26 fall). I nästan lika många fall (LHM = 21) har det skett en felstansning i satskolumnerna. Eftersom detta sker i ett fall per hålkort, kan man i fortsättningen räkna med att .05 % av korten kommer att felstansas i det här avseendet, således i mycket blygsam omfattning.

Återigen visar sig stansfelen till största delen inom textavsnittet, men även kodningsfel förekommer som inte beror på ändringar gjorda i efterhand. Tydligen förbises parenteser i några fall, som dock inte bör förekomma. Allvarliga är också negationsfelen, även om de är få. Att inte koda ett påstående som nekat i det fall det är det, betyder självklart en felaktig analys i fortsättningen. På detta stadium går det dock inte att överblicka konsekvenserna av denna feltyp. Vid ett studium av varje enskilt fall har åtminstone en förklarlig orsak visat sig, nämligen att den ena kodaren missbedömt adverbet 'aldrig' vid ett flertal tillfällen. Begreppet har kodats som tidsadverbial i stället för negationsadverbial, såsom reglerna anger. Felen kan dock förklaras med att intervjuformuläret innehåller skattningsskalor angående utnyttjande av informationsorgan, där polen 'aldrig' kan ses som ett tidsbegrepp.

I det följande redovisas feltyperna C 3 och D 1, som avser stansfel och kodningsfel inom text med tillhörande kod. Eftersom registreringsprinciperna här skiljer sig från principerna vid kontrollstansningen, beräknas inte andelen av antalet nedslag. I stället beräknas den relativa frekvensen dels inom respektive kategori (k), dels per antalet hålkort (h), totalt 36 689.

Tabell 6. Jämförelse mellan stansfel (LHM) och kodningsfel (K) vid en granskning av det totala stansade materialet. Text och innehållskod (observerad och relativ frekvens beräknad på respektive feltyp) och antal hålkort

Feltyp	LHM			K			Tot		
	f	% k	% h	f	% k	% h	f	% k	% h
C 3	7	8.6	.02	74	91.4	.20	81	100	.22
D 1	35	12.9	.09	237	87.1	.64	272	100	.73

Som framgår ur tabell 6 svarar kodarna för omkring 90 % av felen inom både text och innehållskoder. Beräknas felen per hålkort uppgår de till .20 %. I detta sammanhang bör det betonas att det procentuellt ytterst ringa antalet fel skulle, om vi beräknade per nedslag, som skedde vid redovisningen av kontrollstansningen, reduceras ännu mer. Det har redan framhållits att dessa båda kategorier är materialets väsentligaste.

Det procentuellt större antalet kodningsfel kan antas bero på att kodningen skedde under en relativt lång tidsperiod, under vilken kodarnas referens till reglerna kan ha förändrats. Tolkningen av begreppsrelationer (t ex adverb/attribut) kan vackla. Vi befinner oss fortfarande i ett utvecklingsskede, och justeringar gjordes i manualen under kodningen.

Detta kan förklara många om än inte alla fel. Det har inte gått att undvika felskrivningar under D 1, och såväl felskrivningar som feltolkning av text får allvarliga konsekvenser för den senare analysen. Därför kan det vara på sin plats att närmare se på vilka olika fel som uppstått inom kategori D 1. Den speglar samtidigt ändringar inom C 3. Här väljs att beskriva de olika skiftningarna inom innehållskodningen av den anledningen att den numeriska beskrivningen torde vara överskådligare. Först redovisas stansfelen, därefter kodningsfelen.

Kategorin "30, 50, 70" innebär subjekt och objekt, huvudkategorier. Attributen kan tillhöra dessa tre. Skiftningen inom verb gäller kopulaverb i förhållande till övriga verb (kod 41 resp 40). Även loop-koden redovisas inom D 1 (se redogörelsen under ruta 5). Markering inom positionen "stansat" 30, 50, 70 och "kodat" 30, 50, 70 utläses så, att byte av huvudkod har skett. Om ett adverb har bytt kod till attribut, betyder detta samtidigt att det byter huvudkod, eftersom den första siffran i den tvåsiffriga koden ändras från 4 till 3, 5 eller 7 (se ruta 8).

Tabell 7. Felstansning i det totala stansade materialet. Innehållskoder (D 1)

Kodat Stansat	30, 50, 70	Attr.	Adv.	Verb	99	88	++	--	noll	loop	Tot	%
30, 50, 70	1	3									4	11.4
Attribut	1	1									2	5.7
Adverb			2						1		3	8.6
Verb			1	9							13	37.1
99				1	2						3	8.6
88												
++												
--		1	1								2	5.7
noll-stans			1(2)							2	4	11.4
loop-kod										4	4	11.4
Tot	5	5	6	10	2				1	6	35	
%	14.3	14.3	17.1	28.6	5.7				2.9	17.1		100

LHM tot. av D 1 = 12.9 %

Tabell 7 visar att stansfelen är flest inom verbkoden, där huvudkod för verb är 4, vilken i de flesta fall förväxlats med siffran 9, troligen ett resultat av otydlig handstil. Stansfelen upptar dock en blygsam del (ca 13 % eller .09 % av hålkorten) och kräver knappast någon ytterligare kommentar.

Tabell 8. Felkodning i det totala stansade materialet. Innehållskoder (D 1)

Korr. till kod. stans.	30, 50, 70	Attr.	Adv.	Verb	99	88	++	--	noll	loop	Tot	%
30. 50. 70	35	13	5	5		4					62	26. 2
Attribut	12	29	7	2	5	2		2			59	24. 9
Adverb	12	12	32	1				2			59	24. 9
Verb	4	3	1	10	1	1					20	8. 4
99	3		2			3					8	3. 4
88												
++			2		1		1				4	1. 7
--	1		2								3	1. 3
noll-stans	2 (2)		5 (2)					2			16	6. 8
loop-kod										6	6	2. 5
Tot	71	57	61	18	7	10.	1	6		6	237	
%	30. 0	24. 0	25. 7	7. 6	3. 0	4. 2	. 4	2.5		2. 5		100

Tabell 8 visar att det inom kodningsfelen särskilt är tre fall som är anmärkningsvärda. Flest fel förekommer i kategorin huvudord (30, 50, 70). Denna kategori är tillsammans med verben den väsentligaste, och fel här är allvarligare än i övriga fall. Men resultatet var ändå inte helt oväntat, eftersom särskilt tolkningen av de båda objekten (50, 70) stötte på svårigheter under kodningsskedet. Reglerna har måst ses över under tiden, varför vi hade räknat med sådana ändringar.

Därnäst visar sig ändringarna från ett adverb till ett annat vara många. Det rör sig ofta om tolkningar av sätts- och gradadverbial jämfört med orsaks- och omständighetsadverbial (kod 45 resp 46) och ibland även typen satsadverbial kontra tidsadverbial (kod 42 resp 43). Dessa osäkerheter uppmärksammades också under kodningsperioden, men kodningsfel här kan inte betraktas som särdeles grava, då första siffran i koden kvarstår (förväxlade med verb kan de ändå inte bli).

Den tredje stora gruppen är korrigeringar inom attributen. Dessa får också betraktas som allvarliga, främst i de fall en korrekt kodning inte görs vid adjektivattributen, eftersom dessa särskilt skall behandlas för skalering.

Resultatet av den totala kontrollen visar att kontroller av det här slaget är nödvändiga, särskilt när det gäller kodning av komplex natur med både text och siffror. Kontrollen visar också, att trots detta för stansare förhållandevis svåra material sker stansfel i mycket blygsam omfattning. Där de sker gäller de huvudsakligen rena felslag, som en korrekt kodning inte kan förebygga.

7. 3 Sammanfattning och rekommendationer

Kontrollstansningen har visat att LHM har utfört en säkrare stansning än LDC. Detta gäller såväl numeriska som alfabetiska tecken, av vilka de senare generellt är svårare. De flesta fel uppstår inom text och gäller felslag p g a närhet mellan tecken på tangentbordet. Övriga textfel tycks bero på otydlig kodning med avseende på handstil, en feltyp som borde kunna elimineras. De stora skillnaderna i stansningen till LHM fördel skulle till en del kunna förklaras med att stansoperatrisen och kodarna har haft kontakt med varandra personligen, och att hon så småningom har lärt sig känna igen kodarnas skrivsätt, vilket LDC inte haft möjlighet till. Om detta faktum är en väsentligt bidragande orsak till ett gott stansresultat, finns det ingen som helst anledning att lämna bort ett material för stansning utan vi bör se till att stansning kan ske på institutionen.

Kontrollen av hela det stansade materialet visar att inkonsekvenser har förekommit i kodningen, framför allt inom kategorierna huvudord (30, 50, 70), adverb och attribut, av vilka den första är den allvarligaste, men förklaras genom den utveckling regelsystemet undergick samtidigt med kodningen. Inom de viktigaste kategorierna svarar kodarna för ca 87 % av felen, eller .64 % av hålkorten.

Stansningen är mycket säker. Flest fel förekommer inom text (stavning), men eftersom dessa är rena felslag kan man räkna med ett visst antal sådana fel även i fortsättningen, totalt .02 % av nedslagen.

Om kodningen är konsekvent, noggrann och tydlig samt om ändringar i efterhand inte måste företas, kan vi räkna med ett nära nog perfekt material.

För att i största möjliga mån kunna skapa garantier för korrekt stansning bör kodare tänka på följande:

1. Ju fler alternativa koder som är möjliga inom en given kolumn, desto större är risken för stansfel och desto tydligare måste handstilen vara.
2. Stansning av numeriska tecken är lättare än alfabetiska, varför det är särdeles viktigt att handstilen hos kodaren gör klar åtskillnad mellan särskilt följande tecken:

konsonanterna	t/f, k/h, v/r, h/n
vokalerna	a/u
vokal-konsonant	u/v

3. Skriv inte med suddigt stift och radera ändringar.

Den som ansvarar för stansinstruktionen bör inte försumma att:

4. Informera stansoperatrisen mera informellt om materialets karaktär, så att hon själv kan upptäcka inkonsekvenser. Muntliga kontakter mellan operatris och kodare ger mer motivation och ökar förståelsen för bådas arbete, vilket bara kan befrämja slutprodukten.

DEL III. MANUAL FÖR INNEHÅLLSANALYTISK KODNING AV INTERVJUTEXTER

INNEHÅLL

FÖRORD

Denna manual bör betraktas som ett arbetsmaterial som är under utveckling. Den nu föreliggande versionen är en första revision av ett material som utprövades under våren 1974, då vi ville få en uppfattning om dels hur den här typen av kategorisering och kodning kunde tränas in av flera kodare, som inte var involverade i utvecklingen av regelsystemet, dels på vilka punkter svårigheter uppstod.

En första kodning av vissa delar av huvudmaterialet har skett med ledning av manualen i det skick den nu har. Vi är medvetna om svårigheten att täcka alla språkliga varianter med det relativt ringa antal koder som hittills finns. På flera punkter kommer regelsystemet att ändras och utvidgas. Förslag till sådana ändringar gjordes under kodningens gång, men har inte införts, eftersom det skulle försvåra en jämförelse med resultatet från detta kodningsarbete. I de inledande kapitlen skulle en hel del ändringar ha kunnat företas i och med att vår lingvistiska utgångspunkt delvis har ändrat karaktär, men kapitlen kvarstår i oreviderad form för att underlätta senare diskussioner om kodningens resultat.

Till sist skall nämnas att reglerna utarbetats på grundval av ett urval av provintervjuer. Till hjälp för precisering av kategorier har jag haft dels ett manuskript till manual för meningsbyggnadsanalys, som professor Ulf Teleman välvilligt ställt till mitt förfogande under en kort period hösten 1973, dels Olof Thorell, Svensk grammatik från 1973. De exempel som ges till reglerna och i den särskilda exempelsamlingen är till största delen hämtade från urvalet av provintervjuer.

Malmö i juni 1974

Inger Bierschenk

1. KORT ORIENTERING

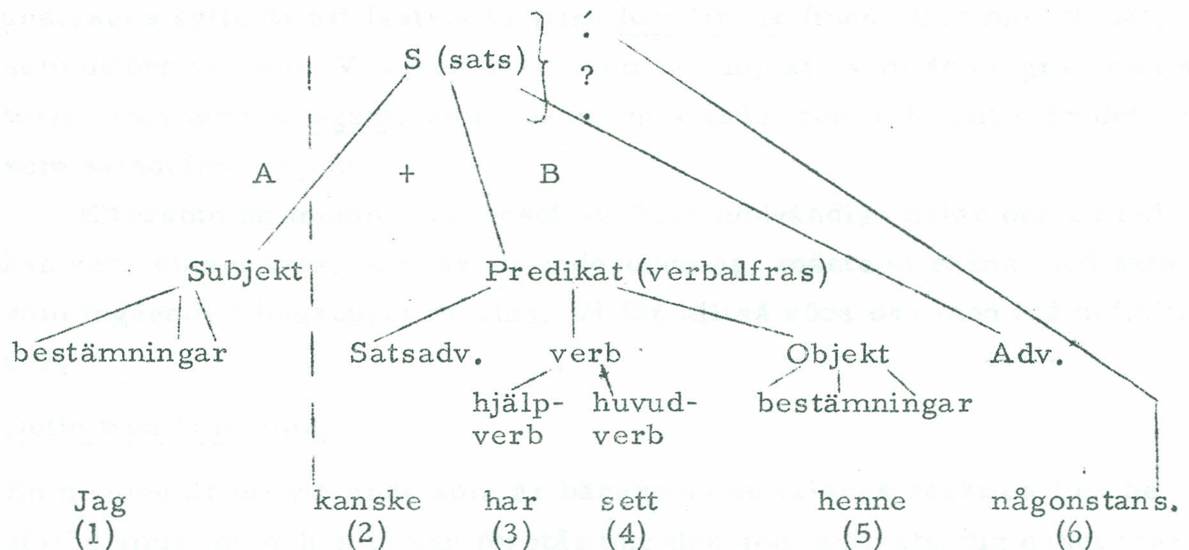
Inom SÖK-projektet vid LHM försöker vi utveckla ett program för datorbaserad innehållsanalys av intervjuer. Insamling av material företogs i maj 1973 och skedde med hjälp av bandspelare. De bandade intervjuerna har under hösten 1973 nedskrivits och textmaterialet omfattar ca 4000 sidor. Dessa skall först prepareras manuellt för att sedan kunna behandlas maskinellt. Regler för en sådan manuell kodning har utarbetats på grundval av ett provmaterial.

Analysen bryter ner texterna i mindre praktiskt bearbetningsbara enheter. Eftersom analysen är baserad på lingvistiska teorier krävs av kodare en god orientering i språkstruktur i allmänhet och svenska språkets uppbyggnad i synnerhet. Även om den språkteori som ligger bakom analysen något skiljer sig från traditionellt sätt att betrakta grammatiken bör kodare i första hand ha goda kunskaper om den "vanliga" satsdelslösningen för att snabbt kunna tillägna sig våra kodningsregler. Kännedom om den s k generativa grammatiken och annan modernare språkteori är en fördel.

2. ANALYSENHETER

Vi analyserar texten med utgångspunkt i den traditionella meningen (sentence), som inkluderar satsbegreppet. För att ett yttrande skall vara meningsfullt skall det vara förståeligt, vilket inte behöver vara detsamma som att det är grammatiskt korrekt. Ett enda ord kan ju vara meningsfullt när det yttras i ett visst sammanhang till en viss person. Men för att vi skall kunna utföra vår analys kräver vi en basenhet: handlingen, verbet. Kring denna bas vill vi bygga upp strukturen i ett yttrande. Vi utgår från att det finns en understruktur, nämligen talarens tanke (vad han vill säga) och det som i realiteten sägs, dvs hörs.

Syntax innebär ordningsföljd eller relation mellan de olika leden i ett yttrande. Kring verbet grupperar sig vissa adverbial, kring substantiv ("subjekt" eller objekt) ordnar sig olika bestämningar. Andra delar, t ex tyngre adverb-led är mera rörliga och kan stå i början eller i slutet av en mening eller sats. Frasstrukturgrammatiken placerar dem efter övriga led. Med ett träd skulle en typsats mycket förenklad se ut så här:



Utifrån denna basstruktur kan sedan samma sats omstruktureras genom omflyttning av leden. Detta kan åskådliggöras genom ett positionsschema:

kanske	har	jag	sett	henne	någonstans
någonstans	"	"	kanske	sett	henne
henne	"	"	"	"	någonstans
sett	henne	någonstans	har	jag	kanske
"	"	har	jag	kanske	någonstans
jag	har	kanske	sett	henne	"
"	kanske	har	"	"	"
någonstans	"	jag	har	sett	henne

Denna positionsanalys visar den bestämda ordningen (syntaxen) i satsen beroende på vilket ord som väljs som "startord" (även kallat fundament). Fler möjliga kombinationer kunde ges utan att samma relationer skulle rubbas. Den ordning som omedelbart framträder visar vilka led som är direkt beroende av varandra: verbet + objektet. Endast i ett fall bryts objektet ut från huvudverbet. Subjektet berör inte omedelbart objektet. Om vi tittar tillbaka på träddiagrammet ser vi att detta uttrycks genom att subjektet är fristående men att predikats- eller verbalfrasen som helhet står lika fristående. Delarna i verbalfrasen är omedelbart beroende av varandra.

Ett sådant resonemang kan nu föras när det gäller att definiera vår analysenhet: meningen. Den traditionella satsanalysen utgår från att en fullständig mening skall innehålla subjekt och predikat, dvs de huvudbeståndsdelar som en sats består av. Vi har inget emot en sådan definition, men eftersom vår basenhet är verbet eller handlingen måste i första hand objektet fastställas som omedelbart berörs av verbet. Därtill kommer att analysens syfte är att fastställa vilka handlingar (med eller mot objekt) som utförs av vem. Vi byter då ut termen subjekt, som är en grammatisk term, mot termen agent, som bekant en satsdel som inte alltid är detsamma som satsdelen subjekt.

Eftersom en mening kan bestå av flera nödvändiga delar och en del kan vara en hel sats, som är en underordning, måste vi räkna med sats som ingående i begreppet mening. Vi får alltså röra oss med två definitioner.

Definition 1: mening

En mening är ett yttrande som är bärare av en talares reaktion i en bestämd situation och som kan förstås när den som självständig enhet lösgöres från sin omgivande kontext.

Definition 2: sats

En sats är en meningsenhet som uttrycker en handling eller ett tillstånd och som kan vara 1) bärare av självständigt innehåll eller 2) bärare av innehåll endast tillsammans med omgivande kontext som den domineras av.

Basenheten handlingen/tillståndet

Vår typmening innehåller handlingen att se (i förfluten tid). Detta innehåller ingen information om vi inte vet vad som ses. Henne kompletterar alltså handlingen se. Hjälpverbet har ger ingen nödvändig information till handlingen. Däremot finns det andra enheter som kompletterar informationen att se henne. Kanske och någonstans uttrycker att handlingen inte

säkert har inträffat, resp. osäkerhet om var handlingen ägde rum. Agenten jag omtalar vem som utförde handlingen.

En handling kan också bestå av t ex att ge. Men ger henne bär inte upp någon information beroende på att verbet giva måste ha ett komplement ytterligare, t ex boken. Ger boken är heller inte tillfredsställande, eftersom handlingen att ge boken saknar målet mot vilket den riktar sig. Ger henne boken är alltså fullständig. Vissa verb måste ha två objekt, ett som omedelbart (direkt) domineras av verbet, ett som är handlingens mål. I traditionell satsanalys säger vi att boken är direkt objekt och henne indirekt.

Vår analys vill ta reda på vilka handlingar (verb, verb + direkt obj.) som riktas mot vilka mål samt vem som agerar. Dessutom vill vi veta "tillstånd" som återspeglas hos agenten eller hos objekten. De uttrycks ofta genom verb men lika ofta genom adjektiv som beskriver någon eller något och som kopplas som bestämning (se satsträdet) eller genom en kopplande länk, som t ex "är" (kopula). Det satsled som kommer efter "är" kallas predikativ och kan inte betraktas som komplement till en handling utan som komplement i form av bestämning eller beskrivning till det satsled det kopplas till, antingen detta led är agent eller objekt.

Självständigt yttrande

Om vi antar att en talare stimuleras till en reaktion och avger ett yttrande, som skall förstås av en åhörare, kan detta yttrande få olika form (representation) beroende på vem som yttrar och vem som hör samt i vilket tillstånd, i vilken situation och med vilket syfte detta sker. Tanken eller reaktionen måste vara en självständig enhet men själva formen kan ha uteslutit vissa delar som bassatsen kan bestå av. För att vår analys skall bli fullständig med avseende på strukturen agent — handling — mål måste vi kunna utläsa dessa delar under ytan av det som sägs. Vi supplerar in (kompletterar) de delar vi är intresserade av, om vi kan utläsa dem under ytan.

Exempel: Jag skickade rapporten. (→ mål)
agent verb dir. obj.

Om vi ur sammanhanget vet målet för handlingen "att skicka rapporten", kompletterar vi med t ex "till SÖ"
mål

Vi blev utsatta för kritik (← agent)
mål verb dir. obj.

t ex "från elever"
agent

När vi således har bestämt vilka analysenheter vi vill ha och vad analysens målsättning är behöver vi samla de regler som bör tas hänsyn till inför en kodning av materialet. Vi måste veta:

1. vilka yttranden i hela textmaterialet som skall analyseras
2. hur vi bearbetar yttranden till relevanta analysenheter samt
3. hur innebörden i analysenheterna kodas på ett praktiskt sätt.

3. ANALYSENS OMFATTNING

Endast intervjupersonernas tal ingår i analysen. Intervjuarens frågor och kommentarer utgår.

Strykningar:

1. Frågor som intervjupersonen ställer till intervjuaren och som rör själva intervjusituationen, t ex ifyllande av skattningsskalor eller dylikt utgår.

   Ex: Kan du ta loss dom här bladen?
   Vilken vill du jag ska besvara först?

2. "Ansatser" till yttranden, dvs sådana meningar som antyder att intervjupersonen tänker svara på en fråga utgår.

   Ex: Få se nu vad jag kan komma ihåg ...
   Jag ska ta ett exempel ...
   Men du, en sak till ...

I vissa fall kan intervjuarens tal flätas in i intervjupersonens. Om man kan utläsa vad intervjupersonen menar, medtages hela yttrandet såsom utsagt av intervjupersonen.

Ex: I: Så att, dom här problemen har du inte ...
Ip: funderat så mycket på, nej.

dvs Ip: Jag har inte funderat så mycket på dom här problemen.
(Se vidare om suppleringar)

3. Vissa strykningar företas vid segmenteringen av materialet (punkterna 1-2). Vid avgränsning av mening kan vi inte alltid utgå ifrån att varje mening avslutats med punkt. En enhet inom två punkter kan bestå av flera meningar, antingen åtskilda genom pauser som i utskriften markerats med flera punkter, eller fragment, som kan suppleras till fullständiga meningar eller satser. Om sådana fragment av yttranden inte kan förstås (vilket inbegriper supplering) utgår de ur analysen.

4. Korrigeringar av talaren

   Ex: Det var projektledaren som först påpekade det ... tror jag ... jo, det var det.

   min professor ... nej förresten det var biträdande som tillstyrkte ansökan.

I första fallet stryks tror jag, eftersom talaren omedelbart efteråt blir säker på sin sak. Ja, det var det är bara en bekräftelse av det sagdas giltighet och medtages inte heller. I det andra fallet utgår min professor nej förresten.

Allmän regel. När talaren korrigerar sig stryks "det som inte gäller".

4. SUPPLERINGAR

Vissa meningar kan vara fragment som vi inte kan supplera till självständiga enheter. I sådana fall är det klokast att låta partiet utgå. Tanken bakom yttrandet skall vara fullt förståelig för kodaren.

Suppleringar sker med parentestecken för att analysen inte skall tappa bort vad intervjupersonen i realiteten säger.

Olika slag av supplering

1. Ja- och nej-svar

Enbart ja, jo, nej eller liknande som svar på en fråga måste suppleras med ett helt påstående, jakat eller nekat. Kodaren måste alltid utgå från vad intervjuaren säger och endast använda hans formulering.

Ex: I: Har du använt dig av personliga kontakter vid informationssökningen?
Ip: Ja.

Utsagan lyder: Ja (jag har använt mig av personliga kontakter), resp. Nej (jag har inte ...)

Observera att svarsorden får särskild kod (99), om svaret fortsätter efter detta ord. (se sid 18)

2. Ersättningsord

En mängd ord syftar tillbaka på något redan sagt. Lösryckt ur kontexten blir yttrandet inte begripligt om sådana relationsord inte får sin betydelse (referens) angiven.

Sådana ord är a) han, det, dom, man, som osv
b) då, nu, där osv
c) sådan, alla, annan osv.

Särskilt fall är jag som inte behöver suppleras, eftersom jag har sin referens angiven i särskild kod. (Se datablanketterna.)

Man, vi eller du kan ibland vara ersättare för jag. Om denna relation är tydlig suppleras (jag) in.

Vi kan vara liktydigt med projektmedlemmar eller enbart projekt eller institution eller någon annan grupp. I de flesta fall talas det om projektarbete. Där vi är liktydigt med projektet suppleras (projektet). Det måste också specificeras vilket projekt det gäller: (XYZ-projektet). (En lista över projekt medföljer separat.) När en handling omtalas, där det klart framgår att enskilda personer inom projektet avses, suppleras (projektmedlemmar).

När någon, annan eller liknande förekommer utan utpekande referens suppleras det med (person). Detta är viktigt för att man inte skall förväxla med någon (+ "sak").

Ex: Där (inom civilförsvarsstyrelsen) arbetade vi (jag) med befattningar.

Den (boken + titel) hade ju många (personer) läst.

Han (NN) var med på konferenserna. Det (NN:s erfarenhet) passade ju bra här (i XYZ-projektet).

Jag blev tjänstledig då (1 juli 1967).

Boken, som (boken) han (NN) skrev, utkom aldrig.

Jag hade en sådan (yrkespedagogisk) bakgrund.

Vi (projektmedlemmarna) diskuterade nya uppgifter.

3. Betydelsefattiga ord

a. verb

Vanligast är verbet göra (jfr eng. 'do'), som måste suppleras med det verb det i sammanhanget ersätter.

Ex: gör (arbetar med)
gör (undersöker)
gjorde (medförde)

Märk: om göra = 'tillverka' behålls det.

I många fall måste 'göra' bli detsamma som sitt objekt.

Ex: Vi har gjort (avgränsat) en avgränsning.

Man måste särskilt uppmärksamma de förskjutningar i betydelsen som förekommer:

Ex: En projektbeskrivning gick in (skickades) till SÖ.
Arbetet ligger på viss tid.

Om 'vara' = befinna sig suppleras det inte.

b. substantiv

Vi (XYZ-projektet) har haft hand om många saker (ex. undersökningar).

4. Ersättningsord till hel sats

Det förekommer ofta att intervjupersonen svarar med ett "det" som omfattar hela innebörden i frågan.

Ex: I: Hur kom det sig att du blev intresserad av detta problemområde?
Ip: Det är svårt att säga.

'Det' syftar på hela frågan, dvs Det (hur det kom sig att jag blev intresserad av detta problemområde) är svårt att säga.

För att undvika sådana långa suppleringar använder vi en standardiserad supplering. Varje fråga har ett eller flera ord som en sammanfattning av sitt innehåll. Genom denna standardisering undviks en alltför olik-

artad kodning, vilket är risken vid kodning genom flera personer. Det är dessutom arbetsbesparande. (Standardiserade begrepp tillhandahålls separat.) Exemplet ovan skulle se ut så här:

Ip: Det (intresset för problemområdet) är svårt att säga.

Eftersom 'problemområdet' är känt suppleras dessutom in t ex projektnamnet eller delundersökningens arbetsnamn, om detta uttrycks explicit av Ip.

5. Satsrelationer

a. att-infinitiv

Ex: Jag bad honom att få komma.

Vi utgår från att satsen har ett verbbegrepp (huvudverb). Här finns två: bad och komma. I sådana fall är agenten underförstådd för det senare verbet och meningen har två satser:

Jag bad honom (om något) resp.

(jag ville) få komma.

Eftersom övervägande delen att-infinitivkonstruktioner använder 'skall' eller 'skulle' supplerar vi in verbet skola i lämplig form.

Ex: Jag bad honom att (jag skulle) få komma.
Jag ber honom att (jag skall) få komma.

b. för att-infinitiv

Ex: Han ringde mig för att få upplysningar.

dvs: Han ringde mig
för att (han skulle) få upplysningar.

Märk: Jag bad honom att komma.

Jag bad honom
att (han skulle) komma.

c. relativ-satser

Ex: Boken, som (boken) utkom igår, har jag inte hunnit läsa.
Ett projekt, som (ett projekt) kallas Skrivsyntax, startades.

Varje mening innehåller två satser:

Boken har jag inte hunnit läsa.
Boken utkom igår.

Ett projekt startades.
Ett projekt kallas Skrivsyntax.

Vidare:

Jag arbetade inom en grupp, där vi hade stor sammanhållning.

Ibland går det ju perioder, då man inte läser någonting.

d. icke-underordnade relationer

Ex: Jag började sent med min projektskiss. Därför blev den inte så utförlig.

Därför, av den anledningen, m fl inleder ny mening men måste suppleras på samma sätt som ersättningsorden. Exemplet ovan blir:

Därför (eftersom jag började sent med min projektskiss) blev den inte så utförlig.

Orsaksrelationen därför ersätts i den underordnade satsen med eftersom + orsakssatsen. På samma sätt ersätts då med när, exakt tidpunkt, medan; sedan med när; därefter med efter (det) att.

Allmän regel

Vid val av ord som skall suppleras används i första hand det som intervjupersonen själv använder, om inte sammanhanget gör det omöjligt.

Ex: Det var mera common sense-attityder och ska vi säga erfarenheter från fältet dom (lärarkandidaterna) appellerade till. Aldrig att det och det läste vi i pedagogik.

(Lärarkandidaterna appellerade) aldrig ...

Observera att materialet nästan uteslutande rör sig om forskning. Först bör man pröva om forskning eller forskare är lämpliga suppleringar (t ex vid ingen, någon etc) innan man använder person. Suppleringarna skall ge mer information än det supplerade ordet.

5. KODNING AV ENHETER

Analysen avgränsar varje enhet som innebär information och som har annan funktion i satsen än övriga delar. Varje enhet, bestående av ett eller flera ord, tilldelas ett kodnummer. Dessa nummer är uppdelade i fyra huvudkategorier 3, 4, 5, 7 (ej 6, då 5 och 6 lätt kan förväxlas vid läsning av datautskrifter). Dessa motsvarar de fyra huvudkategorierna agent, handling, objekt och mål. De underkategorier som fungerar som bestämningar av olika slag tilldelas siffror som omedelbart domineras av huvudkategorin: Agent får t ex siffran 30 (0 = huvudord) och ett adjektiv som beskriver agenten får siffran 32.

Förutom satsens enskilda delar kodas utsagans tempus, modus m fl övergripande strukturer som kodningen av enheterna inte kan ge. Nedan redovisas våra innehållskategorier tillsammans med kodnummer och sådana enheter som för analysen inte är av betydelse, samt ett kodningsförfarande av teknisk karaktär. Därefter ges anvisningar för användningen av datablanketter samt exempel på utskrift (kap. 6).

Innehållskoder

Agent/subjekt	Kod 30
1. substantiv	Per, projektet, SÖ
2. pronomen eller myckenhetsord, där substantiv saknas	Han, jag, vem, två, många, andra, inga
3. självständiga adjektiv eller particip	Det svåraste återstår Den tillförordnade
4. infinitiv	Att bestämma sig var svårt Det var nödvändigt att informera om möjligheterna
5. underordnad sats	Att uppdraget gick till mig kan väl bero på min erfarenhet Om vi kommer är osäkert Hur det sker har inte bestämts Vem som styr kan lätt påvisas

Attributiv bestämning framför huvudordet	31 (51, 71)
1. a) totalitet	alla, hela, inga
b) utpekande	denna, dessa, de (den här)
c) possessiv/genitiv	dina, hans, mitt, tidens
d) kvantitet	många, flera, tusen
e) urval	första, övriga, de återstående, andra
f) jämförelse	sådan, liknande, dylik
2. adjektiv som beskriver myckenhetsord	Ett högt mått av intolerans En hel (stor) hög med ärenden
(uttrycket kan skrivas om till myckenhetsord	en del, ett gäng (+ subst.), mycket, massor av, många, några)

3. adverb som bestämmer myckenhetsord	ganska många, relativt få
Attribut som beskrivning eller värdering	32 (52, 72)
1. adjektiv	Stor, öppen, brittisk, typisk, ny
2. particip	genomgripande, efterlängtad, reviderad
3. adverb i vissa fall	En annorlunda bok En framåt kille Gradvisa förbättringar
4. adverb till beskrivningen	ganska fin, mycket svår (nästan, alldeles, tämligen)
5. bestämningsord som kan suppleras med adjektiv.	sådan (relevant) information
Attributiv bestämning efter huvudordet	33 (53, 73)
1. adverb	Huset därborta. Arbete inomhus. Priserna då. Platsen framför. På planet hem. Dagen innan. Undersökning regionvis.
2. prepositionsfras	Lärarhögskolan i Malmö. Bussarna på linje 3. Metodiken i klassrumsundervisning. Intresset för problemområdet.
3. subjunktionsfras (uttrycker funktion)	Namnet som sådant. Som delägare vill jag inte uttala mig.
4. infinitivfras	Konsten att lyckas i affärer. Idén att starta projektet. Alternativ att välja mellan.
5. attributiv satsförkortning	Han gick, utom sig av oro (ovanligt i talspråk)
6. underordnad sats	Det faktum att vi är här Frågan vilket parti som har rätt Vi ville inventera problem som lärarkandidater ställs inför
7. apposition	Lektorn personligen. Jag själv kom ensam. NN, institutionens vaktmästare Umeå, universitetsstaden i norr
8. framhävande tillägg (ej sats)	som föregås av: bl a, nämligen, det vill säga, till exempel, m fl. En svår tid, nämligen 70-talet Jag plöjde igenom boken, dvs Gulagarkipelagen. (Se bilaga 2.)
9. när framförställda bestämningar inte kan hänföras till 31 (51, 71) eller 32 (52, 72)	
a) epitet	byrådirektör Svensson
b) genitiv som kan göras om till prepositionsuttryck	skolans utbyggnad (= utbyggnaden av skolan)
genitivformen uttrycker inte ägande	Projektets tillkomst Statens psykologisk-pedagogiska bibliotek

Predikativ	30 —— 30
Bestämning till "agent" vid verben vara, bliva, heta, kallas. Även: anses, tyckas, förefalla, verka, se ut, representera m fl.	Projektet blev omfattande. 30 32 Hindren föreföll oöverstigliga. Rapporten blev till ingenting. Hon är inte att lita på. Hustrun fungerade som väckarklocka. Jag satt som sakkunnig i flera år. 33
vara + adj. + particip	Skrivelsen är bra formulerad. 32
En bra regel är att utgå från "balansen".	Uppgiften att etablera kontakter var svår. 30 33 41 32 Uppgiften var att etablera kontakter. 30 41 30
Kopula (länk)	41
verb vid predikativ (enligt ovan)	satt som lektor, är, var, verkar
Verb (tillstånd eller handling)	40
1. finit verbform (presens eller preteritum)	förstår, avslutade, lyckas
2. finit hjälpverb + infinitiv eller supinum	skall överväga, hade diskuterat, skulle ha velat komma
3. infinitiv som kommer efter vissa verb: vilja, kunna, måste, skola, böra, må, månde, få, behöva, törs, tör, torde, lär, verka, komma (att)	Vissa rader bör utgå. Han (NN) kommer att skriva.
Vissa verb kan följas av objekt i infinitiv eller sakobjekt.	Han undvek att skriva. Han undvek telefonkontakter. Jag fick i uppdrag att leda en kurs. 40 50 obj. Jag började med att (jag) försöka etablera 30 obj. sats kontakter. 40 50
4. verbbegreppet kan utvidgas om till verbet hör	
a) preposition eller adverb	söka upp, finnas till, sätta ihop, skriva under
b) substantiv idiomatisk karaktär	hålla hus, komma på kant med, lägga vikt vid, dra fördel av, visa tänder, vinna gehör (för), komma till ro, göra lycka, ha effekt på
c) pronomen (sig) adverb uttrycket kan brytas av ett värderande eller klassificerande ord (här gradadverb)	lägga sig vinn om, ta sig samman, ta för sig, slå bakut hade stor nytta av 40 45 40
d) verb + particip som är en omskrivning för passivt verb	Soldaterna blev sårade i kriget (sårades)

e) även längre uttryck där former av 'vara' ingår:	vara intresserad av, vara insatt i
f) övriga verb + perf. particip:	fått ... ökad ha ... behandlade finns indragen, gå förlorad, etc.
Satsadverbial	42
(negationer	får särskild satskod)
1. a) ett-ordsadverbial som är allmänna, modala, konjunktionella	möjligen, således, kanske, nämligen, också, nog, väl, ju, tyvärr, lyckligtvis
b) särskilda	bara, också (= även), inte ens, redan, ännu, just (just idag), däremot (i går däremot), endast
(kan få särskild satskod)	
2. satsadverbialssats, som har samma funktion som satsadverbial och kan göras om till ett sådant	Han (NN) har medverkat, skulle jag tro (troligen, antagligen) Det är otvivelaktigt att jag anländer i tid. (tvivelsutan, säkerligen) Hon (NN) har många järn i elden. Det är jag säker på. (Se bilaga 1)
Övriga adverbial	43-45
1. Tid. När? Under vilken tid eller skede?	(43) I morgon. Vid årsskiftet. Dagligen. Under problemformuleringsfasen.
2. Rum. Var? På vilken plats?	(44) I Stockholm. Hemma. På skolan.
3. Sätt/grad. Hur? På vilket sätt? I vilken utsträckning/grad?	(45) Alla har lyckats bra. Med vänster hand. (På ett) smart (sätt).
Uttryck (satsinledare) som anger:	46
1. villkor	om, ifall
2. medgivande	fast, även om, m fl.
3. följd eller avsikt	så, så att, för, för att
4. disjunktion	antingen-eller, visserligen-men, å ena sidan- å andra sidan
5. orsak, omständighet	eftersom, därför att (+ sats), på grund av (+ huvudord)
6. komparation	lika --- som, mer --- än, såsom, liksom, ju (--- desto), ännu --- än, --- än, --- som
Objekt	50
Har samma slag av huvudord som agenten/subjektet	
Märk infinitiver (ej modalverb)	De fortsatte att träffas Han fick för sig att resa
Satser	De visste inte hur de skulle bete sig.

Fungerar på samma sätt som predikativ, dvs dom upplevde att avgränsningen var ett problem	Avgränsningen upplevde dom som ett problem (se kod 88, sid 63)
Bestämningar enligt samma princip som för agenten/subjektet	51-53
Mål för handling	70
Ett objekt som anger åt, till eller för vem något sker (positivt eller negativt).	Röda korset hjälpte många utsvultna. (många: 71, utsvultna: 70)
Detta objekt fordrar vanligen att det finns två objekt. Vanliga verb med två objekt är ge, lova, meddela, sända, fråntaga, vägra. Direkt objekt behöver inte vara utsagt.)	En grov projektbeskrivning skickades in till SÖ. (projektbeskrivning: 50, till SÖ: 70)
Bestämningen enligt samma princip som för agenten/subjektet	71-73
Övriga markeringar	++
a. När ett 'och' eller liktydigt uttryck binder ihop två objekt används plustecken	Jag skrev rapporter och PM (och: ++, rapporter och PM: 50)
b. Negationsordet utmärks med minustecken	--
c. Negation och bindning. Hela satsen är negerad.	Jag hade varken lust eller ork. (lust: 50, varken … eller: ++, ork: 50)
Satskoder (tema)	Markeras med siffra i särskilda kolumner (se bil 3)

Innehållskoderna ovan är inte tillräckliga för att karakterisera hela yttrandet. Särskilda koder för satsens art fordras.

	Kolumn		Kod
1. Källan till utsagan	11	Talaren själv anser, tycker, etc	1
		Talaren refererar vad någon annan anser, tycker, etc	2
2. Negation (gäller ej vid negerande verb)	12	Helt negerande (inte, icke, ej, aldrig)	1
		Negerande substantiv (ingen, ingenting) Osäkert negerande (knappast m fl)	2
		Varken (vare sig) --- eller, utan att	3
3. Tempus, uttrycker verbets betydelse i tiden	13	Det som händer eller är i nuvarande tid·	1
		Det som hände eller var i förflutenm tid	2
		Det som skall hända eller vara i kommande tid:	3

	Det bör särskilt uppmärksammas att vissa verb till formen är presens men till funktionen uttrycker förfluten eller kommande tid.		... och när jag kommer dit, så <u>får</u> (2) jag uppleva detta (suppl). Hon (NN) <u>blir</u> (3) nog docent. Problem som är <u>upplevda</u> (2) i klassen. (har upplevts)	
4.	<u>Modus,</u> modifierar huvudverbets betydelse. (Önskan/förmodan = även i kol 22-23.)	14	<u>Indikativ</u>, som uttrycker att en handling eller ett skeende framställs som verkligt, utan att det uttrycker talarens vilja, önskan, förmåga eller liknande	1
			<u>Imperativ</u>, uttrycker en uppmaning som i skrift utmärkes med utropstecken (!).	2
			Konjunktiv uttrycker något ickeverkligt, särskilt vid verbet vara (vore).	3
	Huvudverbet har infinitivform.		<u>Modala hjälpverb</u>. Ex: Jag <u>kan</u>, <u>behöver</u>, <u>vågar</u>, <u>försöker</u> inte. Det <u>ska</u>, <u>börjar</u>, <u>fortsätter</u>, <u>slutar</u>, <u>brukar</u> regna.	4
5.	<u>Villkor,</u> uttrycks vanligen med om ... så	15	Villkorssatsen: Om jag får fast anställning ...	1
			Följdsatsen: så ordnar sig ekonomin (vore jag nöjd)	2
			Arbetet hade förmodligen inte kommit igång, om vi inte haft lämplig personal	(2) (1)
6.	<u>Orsak</u>	16	Orsakssatsen: En orsak till att man valde mig ...	2
			Följdsatsen: var att jag hade rätt ålder	1
	Satskod sätts endast vid fullständig sats. I annat fall täcks uttrycket genom adverbialkod (46)		Jag var med där (suppl) eftersom jag hade kunskaper På grund av arbetslösheten går jag mest omkring hemma.	(2) (1) (46)
7.	<u>Medgivande</u>	17	Medgivandet: Det hela avlöpte väl ...	1
			Inskränkningen: Även om det såg illa ut	2
			Trots vissa svårigheter kunde arbetet genomföras	(2) (1)
8.	<u>Följd/avsikt</u>	18	Försatsen: Hon startade det hela ...	1
			Följd/avsiktssatsen: för att (hon ville) se om det skulle lyckas.	2
			Vi anlände tidigt så (att) vi skulle hinna med.	(1) (2)

9. Disjunktion, kodas där båda leden förekommer	19	Antingen (visserligen, å ena sidan) --- eller (men, å andra sidan)	1
10. Komparation, kodas för förekomst i hela meningen	20	Han (NN) gjorde en bättre uppsats (1) än vad jag lyckades med (46) Somliga har lika hög lön (1) som andra (46) Ju hårdare oljeländerna ransonerar (1) desto kallare blir det (1)	1
11. Indirekt fråga, kodas för förekomst i satsen. Detsamma gäller direkt fråga (som inte är ställd till intervjuaren)	21	Jag ville veta om jag kunde resa Frågan gällde huruvida anslag (1) skulle beviljas Vi har rader av designer att välja mellan. Vilken av dem (suppl) är bäst? (Underförst. undrade vi) (1)	1
12. Antagande, kodas för förekomst (se utskrift på datablad)	22	Den (suppl) skulle förmodligen inneburit en stor omstrukturering Vi (forskare) hade måst pröva frågan (suppl) på nytt, antar jag (skulle jag tro, etc) (1) Jag antar att institutionerna är öppna för flera förslag (suppl) (1)	1
13. Önskan, kodas för förekomst (se utskrift på datablad)	23	Uttrycks på liknande sätt som antaganden genom verb eller adverbial (jag hoppas, förhoppningsvis, önskar, vill)	1

"Innehållslösa enheter" 99

a. Talspråkets småord

Många av de s k ersättningsorden eller andra småord används ofta i talat språk utan att ha egentlig funktion. Man kan naturligtvis diskutera deras eventuella innehåll, men vi menar att de inte ur innehållslig synpunkt är intressanta för vår analys.

Ex: Februarinumret i år handlar om Principals and Vice Principals. Då har dom samlat allt som är skrivet. (då uttrycker ingen tidsrelation.)

Nu gick det inte bra för den där utredningen. (nu anger ingen tid.)

Sen vill jag gärna understryka ... (sen uttrycker endast att talaren fortsätter).

Efteråt, så blev vi inbjudna till konferenser. (så kan vara ett talspråkligt uttryckssätt eller insätts för att göra texten mera lättflytande.)

Vi vill inte stryka dessa småord ur materialet, dels därför att många av dem skulle kunna ha betydelse vid närmare granskning (det finns risk för missbedömning), dels därför att vi så litet som möjligt vill göra intrång i intervjupersonernas sätt att uttrycka sig. Finns dessa småord med i utskrif-

ten tvingas kodaren tänka igenom betydelsen av dem och tilldela dem ett kodnummer. Kod 99 kan sägas vara en kontroll på att vissa "osäkra" ord har kommit med. Det är lätt att efter kodningen kontrollera om det i den "innehållslösa" koden finns sådana ord som kan kodas om, eftersom de kan identifieras.

b. Formellt subjekt/emfas

Ex: 1. Det fanns många projekt på gång.

2. Det var flera (projekt) som hade lämnat in ansökan om tilläggsanslag.

dvs: 1. Många projekt fanns på gång.

2. Flera (projekt) hade lämnat in ...

I kod 99 faller 1. det och 2. Det, var, som.

c. Opersonlig konstruktion

Ex: Det skrivs här
99 40

Det regnar
40

Det skrivs promemorior
40 50

Det är utan agent-innebörd.

d. Samordnande konjunktion

Och, men, för, utan, ty, så, fast länkar överordnade satser med varandra. Deras funktion är inte nödvändig för analysen, men de stryks inte, utan finns med på detta sätt.

Ex: Vi har jobbat väldigt mycket som ett team och vi har inte gjort en klar avgränsning från början.

1. Vi (suppl) har jobbat ...

2. Och vi (suppl) har inte ...
99

e. Ja, jo, nej-svar

Om svaret fortsätter efter ja, jo, nej, m fl förs svarsorden till denna mening som första ord.

Ex: I: Du är handledare?
Ip: Ja. (jag är handledare)
99
I: Har du ett eller flera projekt?
Ip: Ja, man kan ju säga att det är flera, eftersom det består av
99
tre väl avgränsade delar.

I det sista exemplet är frågan så ställd att ja-nej-svar inte är tänkbart. Det är ändå mycket vanligt att ett 'ja' föregår svaret och ett sådant 'ja' får likaså kod 99.

Teknisk kod 88

a. Flätning

Flätning kallar vi samtliga fall där två satser har flätats in i varandra till en sats. Kod 88 tilldelas det ord som måste "brytas ut" och ingå i ny sats, t ex vid emfas.

Ex: Pedagogiken ansåg dom vara intressant.

'Pedagogiken vara intressant' tillhör en sats som är bruten genom 'ansåg dom', som utgör en sats till följd av att talaren själv inte är källan till utsagan. Vi använder 'att' för att kunna "bryta ut" utsagan i två satser:

Pedagogiken ansåg dom (att) pedagogiken vara intressant
88 objektsats

(Se exempelsamlingens sista blad.)

jfr: Pedagogiken ansåg jag vara intressant
30 42 41 32

En utsaga som görs av ip får källkod 1 och en sats.

6. PRAKTISKA ANVISNINGAR OCH EXEMPELSAMLING

Datablankett

För utskrift av text tillsammans med kodmarkeringar används en blankett i A4-format (datablankett), försedd med 80 kolumner, dvs samma antal som hålkortet, på vilket materialet stansas. Varje kolumn kan markeras med ett värde från 0 till 9. Kolumn som i vissa fall hålles blank har därmed också en kod.

Längst till vänster finns identifikationskoder samt koder för källa och satstyp. Därefter följer två kolumner som används för att markera början och slut på en ordenhet som måste sträcka sig över flera rader än en. De flesta kolumnerna upptas av texten, följd av innehållskoder.

Kolumnerna 26-65, text

Texten skrivs ut med en bokstav i varje kolumn samt med tom kolumn efter ordslut. Vid suppleringar anges parentesen i stället för tom kolumn. Accenttecken gäller för bokstav och skrivs i egen kolumn (t ex ide ´). Siffror skrivs med bokstäver utom årtal. Endast vanliga förkortningar får användas (dvs, tex) som skrivs utan mellanrum. (Om förkortningen bildar ett ord med innehåll (tom) skrivs den med mellanrum.) Förkortningar på institutioner används där de förekommer i texten (LHM) och där hela ordet används skrivs det ut så (Lärarhögskolan i Malmö). Interpunktion används inte. Bindestreck av typen låg- eller mellanstadielärare måste delas upp i "lågstadielärare eller mellanstadielärare".

På samma vågräta rad skrivs en ordenhet (ett eller flera ord) som tilldelas en innehållskod som inte är densamma som för enheten före eller efter.

Undantag:

1. Om två på varandra följande enheter var för sig kan bilda en enhet (oberoende av den andra) skrivs de på var sin rad.
2. Uppräkningar som kan skiljas åt genom ´och´, ´eller´ eller kommatecken skrivs på var sin rad även om de har samma kod.
3. En enhet kan brytas genom en annan och sedan fortsätta.

Ex:	1.	Han	30
		satt	40
		på en väldig stol	44
		i sitt arbetsrum	44
	2.	Jag	30
		sysslade med	40
		lågstadielärare	50
		eller	++
		mellanstadielärare	50

3. började 40
   jag 30
   med 40
   projektskissen 50

Kolumnerna 24-25

I de fall där en ordenhet inte får plats på en rad fortsätter den på nästa. Då anges, beroende på hur många rader enheten tar i anspråk, i kolumn 24 antalet rader som innehåller samma ordenhet (kod) och i kolumn 25 ordenheternas inbördes ordning. Om en enhet löper över två rader markeras detta:

24	25		
2	1	text ...	← den första raden av två
2	2	text ...	

Då enheten endast upptar en rad lämnas kolumnerna 24-25 tomma. Avstavning får ej förekomma.

Kolumnerna 66-80, satser

Varje sats har tre kolumner, de första (kolumn 66-68) är huvudsats. Satsernas två första kolumner används till innehållskoderna, den tredje till en hänvisningskod, dvs siffran anger i vilken satskolumn en menings nästa sats följer. Siffran kan alltså visa framåt eller tillbaka till en föregående sats. Det kan nämligen inträffa att ett ord i huvudsatsen först har en bestämning i form av bisats och sedan fortsätter som huvudsats och därefter har ytterligare en bisats, som måste föras ut i närmast tomma kolumn. Principen med satskolumnerna är att varje satskolumn endast får innehålla ett huvudverb.

*

I det följande visas ett antal exempel på kodningar tillsammans med kommentarer. Meningarna har tagits ut ur provmaterialet som typexempel och representerar därför inte en intervjupersons löpande tal. Identifikationskoderna i kolumnerna 1-10 ges därför som exempel en gång och utelämnas i fortsättningen, om det inte finns särskild anledning att införa dem för att betona eller påminna om ett kodningsförfarande.

* De följande sidorna 66-76 kan tyvärr inte reproduceras. Läsaren hänvisas till den manual som utkom 1977:

Bierschenk, I. (1977). Datorbaserad innehållsanalys: Kodningsmanual. Lärarhögskolan i Malmö: Pedagogisk dokumentation, 52.

2013-06-22 / Inger Bierschenk

Bilagor

Bil 1

Lista över adverb som kan vara satsadverbial[1)]

absolut
antagligen
avgjort
beklagligtvis
definitivt
dessvärre
egentligen
eventuellt
faktiskt
för all del
förhoppningsvis
förmodligen
förstås
förvisso
givetvis
gudskelov
gunås
gärna (för mig)
helt enkelt
i grund och botten
i onödan
i själva verket
ju
kanhända
kanske
liksom
lyckligtvis
minsann
måhända
märkligt nog
möjligen
möjligtvis
naturligtvis
nog
nästan
närmast
nödvändigtvis
ofrånkomligen
onödigtvis
otvivelaktigt
oåterkalleligen
relativt
rimligen
rimligtvis
sannolikt
sannerligen
sic
självfallet
självklart
snarast
strängt taget
säkerligen
säkert
troligen
troligtvis
turligt nog
tvivelsutan
tydligen
tydligtvis
tyvärr
törhända
underligt nog
uppenbarligen
utan tvivel
verkligen
visst
väl
äntligen

Satsadverbialssats ("jag-anföring")

Jag tycker att, tycker jag, menar, anser jag (m fl), såvitt jag förstår, det visar sig ju att, förmodar jag, jag förmodar att, jag tror att, jag tror jag får säga att, tror jag man kan säga, skulle jag tro, man kunde väl säga det att, sen var det väl så att, jag får nog säga att, det är möjligt att, m fl.

Många av dessa adverb-uttryck kan också få särskild satskod om de uttrycker förmodan/antagande (kol. 22).

1) Listan hämtad ur manuskript, tillhandahållet av professor Ulf Teleman, forskningsgruppen för Skrivsyntax i Lund. (Teleman kallar dessa adverb för "talarattitydsadverbial".)

Attribut (här adverb) som kan vara framhävande:

bara
bl a
blott
dvs (ej framför hel sats)
enbart
endast
ens
exempelvis
framför allt
företrädesvis
förslagsvis
förrän ([E-E] ---, ej framför sats)
först 'inte förrän' (först igår)
först och främst
huvudsakligen
i synnerhet
inte minst
just
något av
nämligen (framför uppräkning)
precis
redan
såsom (framför uppräkning)
särskilt
t ex
till och med 'even'
väsentligen
åtminstone

[1] Listan hämtad ur manuskript, tillhandahållet av professor Ulf Teleman, forskningsgruppen för Skrivsyntax i Lund.

Bil 3:1

KODNINGSSCHEMA

Kol	Variabel	Pos	Anm
	Identifikation		
1-2	Intervjuperson nr	01-40	
3-4	Fråga nr	01-n	
5-8	Mening nr	0001-n	
9-10	Ordenhets nr inom mening	01-n	
	Sats(tema)koder		se tillägg
11	Källa till utsagan	1, 2	
12	Negation	1, 2, 3	
13	Tempus	1, 2, 3	
14	Modus	1, 2, 3, 4	
15	Villkor	1, 2	
16	Orsak	1, 2	
17	Medgivande	1, 2	
18	Följd eller avsikt	1, 2	
19	Disjunktion	1, 2	
20	Komparation	1	
21	Fråga	1	
22	Antagande	1	
23	Önskan	1	
24	Antal kort (rader) av ordenhet, där enheten inte fått plats på en rad	2-9	
25	Inbördes ordning hos kort (rad) som innehåller samma ordenhet)	1-9	
26-65	**Text**		
	Huvud-bisatskoder		
66-67	Huvudsatskolumner		innehållskoder
68	Hänvisningskolumn till bisatskolumner	2, 3, 4, 5	
69-70	Första-bisatskolumner		innehållskoder
71	Hänvisningskolumn till andra satskolumner	1, 3, 4, 5	
72-73	Andra bisatskolumner		innehållskoder
74	Hänvisningskolumn till andra satskolumner	1, 2, 4, 5	
75-76	Tredje bisatskolumner		innehållskoder

Bil 3:2

77	Hänvisningskolumn till andra satskolumner	1, 2, 3, 5	innehållskoder
78-79	Fjärde bisatskolumner		innehållskoder
80	Hänvisningskolumn till andra satskolumner	1, 2, 3, 4	

Tillägg

Satskoder

Kol

11	Källa till utsagan	1 = anger den talande själv 2 = anger att annan än talaren står för utsagan
12	Negation	1 = inte, icke, ej, aldrig, ingen, ingenting 2 = knappast o. likn. 3 = varken (vare sig) - eller
13	Tempus	1 = nuvarande tid, gäller när utsagan görs 2 = förfluten tid från tillfället då utsagan görs 3 = kommande tid från tillfället då utsagan görs
14	Modus	1 = indikativ 2 = imperativ 3 = konjunktiv 4 = modala hjälpverb
15	Villkor	1 = villkorssatsen 2 = följdsatsen
16	Orsak	1 = orsakssatsen 2 = följdsatsen
17	Medgivande	1 = medgivandet 2 = inskränkningen
18	Följd/avsikt	1 = församtsen 2 = följd/avsiktssatsen
19	Disjunktion	1 = antingen, visserligen å ena sidan 2 = --- eller, --- men, å andra sidan
20	Komparation	1 = förekomst
21	Fråga	1 = förekomst
22	Antagande	1 = förekomst
23	Önskan	1 = förekomst

Översikt: huvudkategorier

Kod	Innehåll
30	agent/subjekt
31	bestämning till agent (framför)
32	beskrivning " "
33	bestämning " " (efter)
40	verb, verbfras
41	kopula-verb
42	satsadverbial (satsmodifierare)
43	tidsadverbial (när?)
44	rumsadverbial (var?)
45	sätts-gradadverbial (hur?)
46	adv. el. andra ord som anger komparation, motsats el. dyl.
50	objekt
51	bestämning till obj. (framför)
52	beskrivning " "
53	bestämning " " (efter)
70	mål för handling
71	bestämning till mål (framför)
72	beskrivning " "
73	bestämning " " (efter)

Översikt: innehåll ur exempelsamling (ss 66-76)

Fall	datablad nr
Andrahandskälla	1, 7, 11
Antagande	2, 4
Attsats	1, 8
Brytning av kod vid ny sats	5
Disjunktion	10
Flera rader	4
Flätning	10, 11
Formellt subjekt	6
Fråga	6, 7
Hänvisning	1
Identifikation	1
Komparation	9
Medgivande	9
Modalt verb	3, 5, 10
Negation	2, 3, 6, 8
Ordnummer	1, 4
Orsak	5
Passivt verb	3
Relativsats	1, 8
Sammanbindning	3, 5
Satsförkortning	5
Satsinledande frågeord	7
Supplering	1, 4, 5, 8
Tempus	1, 9
Verbaluttryck	2, 3, 8
Villkor	9
Önskan	5

8. REFERENSER

Berg, M. Reliabilitetsprövning av en metod för innehållsanalys. Stencil. Malmö: Lärarhögskolan, februari 1974. (a)

Berg, M. Reliabilitetsprövning av en metod för innehållsanalys av intervjutext. Testkonstruktion och testdata, Nr 26, 1974. (b)

Bierschenk, B. Perception, strukturering och precisering av pedagogiska och psykologiska forskningsproblem på pedagogiska institutioner i Sverige. Pedagogisk-psykologiska problem, Nr 254, 1974. (a)

Bierschenk, B. A computer-based content analysis of interview data: Some problems in the construction and application of coding rules. Didakometry, No. 45, 1974. (b)

Bierschenk, I. Manual för innehållsanalytisk kodning av intervjutexter. Stencil. Malmö: Lärarhögskolan, mars 1974. (a)

Bierschenk, I. Manual för innehållsanalytisk kodning av intervjutexter. Stencil. Malmö: Lärarhögskolan, juni 1974. (b)

Bierschenk, I. Erfarenheter från kodarträning och provkodning. Stencil. Malmö: Lärarhögskolan, augusti 1974. (c)

Bierschenk, I. Kontrollstansning och överensstämmelsekontroller av ett kodat och stansat intervjumaterial. Stencil. Malmö: Lärarhögskolan, november 1974. (d)

Brodda, B. (K)overta kasus i svenskan. Papers from the Institute of Linguistics University of Stockholm (PILUS). Stockholm: Institute of Linguistics, Nr 18, 1973.

Fillmore, C.J. The case for case. I: Bach, E., & Harms, R.T. Universals in linguistic theory. New York: Holt, Rinehart & Winston, 1968.

Halliday, M.A.K. Language structure and language function. I: Lyons, J. (Ed.) New Horizons in linguistics. London: Penguin Books, 1970.

Holsti, O.R. A computer content-analysis program for analysing attitudes: The measurement of qualities and performance. I: Holsti, O.R. Content analysis for the social sciences and humanities. Reading, Mass.: Addison-Wesley, 1969.

Osgood, Ch.E., Saporta, S. & Nunnally, J.C. Evaluative assertion analysis. Litera, 1956, 3, 47-102.

Pool, I. Trends in content analysis today: A summary. I: Saporta, S. (Ed.) Psycholinguistics: A book of readings. New York: Holt, Rinehart & Winston, 1961.

Schank, R.C. Conceptual dependency: A theory of natural language understanding. Cognitive Psychology, 1972, 3 (4), 552-631.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: Mc Graw-Hill, 1956.

Tables of the cumulative binomial distribution. Cambridge, Mass.: Harvard University Press, 1955.

1977-11-01